Web Strategies for Fact-Checkers
(and other people who care about facts)

by Michael Caulfield
Released under a CC BY 4.0 License
What Is the Goal of This Curriculum?

The student-fact checking project of the Digital Polarization Initiative (Digipo) is meant to teach students basic web literacies necessary to civic discourse on the web. Unlike many previous attempts to teach “information literacy”, the Digipo project attempts to teach ways of approaching information that are particular to the problems of evaluating information in a decentralized network like the World Wide Web. It also aims to give students web-native strategies and tools for dealing with that environment.

While many information literacy and digital literacy programs have focused on a long series of questions one should ask of content, our approach revolves around giving students web-native strategies. The web, after all, is an interactive medium, and what we have found is that the most adept fact-checkers think through the use of the tools available to them. Strategies connect specific skills to the sorts of concerns readers need to address.

So, without further ado, let's get to the strategies!
Strategy One: Go Upstream

Here’s one of our most important strategies for checking claims on the web: Go Upstream.

What do we mean by this?

Usually, when you hit a web page that makes a claim it will be something like this, from conservative site *The Blaze*:

So the headline claim here true?

Traditional information literacy tells you to start checking this article for credibility. Who is the author? What is the site? When was it last revised?

We’ll do some of that, eventually. But it would be ridiculous to do it on this page. Why? Because like most news pages on the Web, this page actually provides no original information. It’s just a rewrite of another page upstream from it. We see the indication of that here:
Controversial tweets from the U.S. Office of Government Ethics that praised President-elect Donald Trump were approved by Director Walter M. Shaub personally, the Daily Dot reported Friday.

Through a Freedom of Information Act, the Daily Dot reported that Shaub sent an email ordering an OGE official to post the tweets. The series of tweets posted Nov. 30 applauded Trump for his supposed efforts to alleviate conflicts of interest with his businesses.

All the information here has been collected, fact-checked, and written up by The Daily Dot. It's what we call "reporting on reporting". Evaluating this page is meaningless.

So what do we do? Our first step is to go upstream. Go to the original story, so that we can look at the source of that. Our big understanding here is that most of the stuff on the site wasn’t produced by the site, and to evaluate we need to trace where it came from as a first step. In this case that source would be The Daily Dot. When you get to the Daily Dot, then you can start asking questions about the site or the source. And it may be that for some of the information in their article you’d go a step further back, and look at primary sources they reference.

This warning to “go upstream” before you evaluate is particularly important with sponsored content. For instance, a lot of time on a site you’ll see “headlines” like these, which I pulled from a highly regarded technology magazine:
So are lawmakers really concerned about this amazing military/tactical scope? Concerned because it’s just so good? Maybe. But it’s worth noting that this is not a claim that Network World is making, but rather a claim that -- wait for it -- ZeroTac Tactical Scope is making. It’s an ad, served from another site into this page in a way that makes it look like a story.

Sponsored content isn’t always purely an advertisement. Sometimes it provides helpful information. This piece below, for example, is an in-depth look at some current industry trends in information technology, and might be worth reading to the right person.
But when you see that “sponsored” on top of it, you should immediately realize that evaluating *InfoWorld* as a source will tell you almost nothing about the reliability of this article. The source of this article is not *InfoWorld*, but the technology company Hewlett Packard, written by a VP of Hewlett Packard, with no *InfoWorld* oversight.

You can see how this is not just an issue with political news, but will be an issue in your professional life as well. If you go to work in a technology field and portray this article to your boss as “something I read on *InfoWorld*”, you’re doing a grave disservice to your company. Portraying a vendor-biased take as neutral *InfoWorld* take is a major mistake, which could have unfortunate consequences.

Another place you see upstream issues is syndication. Syndication is a method where a third party is able to publish material directly on a site. So, for instance, you can look at the New York Times web page here:
What you'll see is a set of stories on the left ("Germany’s Latest Best Seller", "Ilsis Claims Responsibility") that are written by New York Times staff, but also a thin column of stories ("UK Stock Market Hits Record") which are identified as being from the Associated Press.

And when you click through you'll find a page on the New York Times site, but it is not by the New York Times:

**UK Stock Market Hits Record as Manufacturers Win Business**

LONDON — Britain’s main stock index hit a record closing high Tuesday as a survey showed manufacturers gaining business from the slide in the value of the pound since the country’s decision in June to leave the European Union.

The FTSE 100 index ended the day 0.5 percent higher at 7,177.89, slightly down on its earlier all-time high of 7,205.21.

One of the main reasons why the FTSE 100 has rallied in recent months
Notice that if you are going to evaluate the source of this article that evaluation is going to have very little to do with the New York Times. You’re going to look at the reporting record of the Associated Press.

People get this wrong all the time. One thing that happens occasionally is an article critical of a certain politician or policy suddenly disappears from the New York Times site, and people claim it’s a plot to rewrite the past. “Conspiracy!” they say. “They’re burying information!” they say. A ZOMG-level freakout follows.

It always turns out that the article that disappeared is a syndicated article. AP articles, for example, are displayed on the site for a few weeks, then “roll off” and disappear from the site. Why? Because the New York Times only pays the Associated Press to show them on the site for a few weeks.

You’ll also occasionally see people complaining about a story from the New York Times, claiming it shows a New York “liberal bias” only to find the story was not even written by the New York Times, but by the AP, or Reuters, or some other syndicator.

Going upstream means following a piece of content to its true source, and beginning your analysis from there rather than from the page you happen to be looking at. Your first question when looking at a claim on a page should be “Where did this come from, and who produced it?” The answer is quite often something that has very little to do with the website you are looking at.

Tracking the Source of Viral Content

The examples we have shown so far have been fairly good at providing information about the source of the content. The Blaze story, for example, clearly links to the Daily Dot piece so that anyone reading their summary is one click away from confirming it with the source. The New York Times makes apparent that the syndicated content is from the Associated Press, so that if you wanted to check the credibility of the source you could easily do that.

This is good internet citizenship. Articles on the web that are repurposing someone else’s information or artifacts should make it clear where those articles come from, and, if appropriate, add a link to the original. This is important to creators, because they deserve credit for their work. But it’s also important to readers, who need to check the credibility of the original sources.

Unfortunately, most citizens of the web are not good citizens. This is particularly true with so-called “viral” content on the web -- material that spreads very quickly through hundreds or thousands of people individually sharing it.
When that information pings around a network, very often people decide not to link, or hide the source all together. For example, here is an interesting claim that two million bikers are going to show up for the inauguration. Whatever your political persuasion, that would be a pretty amazing thing to see.

But the source of the information -- Right Alerts Polls -- is not linked.

So here’s where we show our first trick. Select the text “Right Alerts Polls”. Then right-click your mouse (control-click on a Mac), and choose the option to search Google for the highlighted phrase.
It will execute a search for “Right Alerts Polls”. (Remember this right-click action, though -- it’s going to be the foundation of a lot of stuff we do.)

To find the story, add “bikers” to the end of the search:

![Search query](image)

We find our upstream article right at the top. Clicking through, however, we find that this article still doesn’t tell us where the information is coming from either. However, it does have an extended quote from one of the Two Million Bikers organizers:
So we just repeat our technique here, and select a bit of text from the quote and right-click/control-click. What we want is to figure out where this quote came from, and searching on this small but unique piece of it should bring it close to the top of the Google results.

When we do that we see that there are dozens of articles covering this story, using the the same quote and sometimes even the same headline. But one of those results is the actual Facebook page for the event, and if we want a sense of how many people are committing, then this is a place to start.
This also introduces us to another helpful practice -- when scanning Google results (or Bing results, or DuckDuckGo results) novices scan the titles. Pros scan the URLs beneath the titles, looking for clues as to which sources are best. (Be a pro!)

So we go to the Two Million Biker Facebook event page, and take a look. How close are they to getting two million bikers to commit to this?
Well...it looks like about 1,800. That's nothing to sneer at -- organizing is hard, and people have lives to attend to. Getting people to give up time for political activity is tough. But it's pretty short of the “two million bikers” most of these articles were telling us were going to show up.

When we get into how to rate articles on the DigiPo site as true or false, likely or unlikely, we’ll talk a bit about how to write up the evaluation of this claim. My sense is the rating here is either “Mostly False” or “Unlikely” -- there are people planning to go, that's true, but the importance of the story was based around the scale of attendance, and all indications seem to be that attendance is shaping up to be about a tenth of one percent (0.1%) of what the other articles promised.

But the main point here is we would have learned none of this had we decided to evaluate the original page. We learned this by going upstream.

Tracking the Source of Viral Photo

One of the more viral types of content on the World Wide Web is photography. It is also some of the most difficult to track upstream to a source. Here's a picture that showed up in my stream the other day:
OK, so what’s the story here? To get more information, I pull the textual information off the image and throw it in a Google search:

```
shopper upset over double-parked car abc action news
```

About 2,830,000 results (0.57 seconds)

**Shopper upset over double-parked car - YouTube**
https://www.youtube.com/watch?v=9QU7zqdr7fY
2 days ago - Someone didn't appreciate this bad parking job in Oregon, and decided to box it in with shopping carts... The ABC Action News app brings you the latest trusted news and information. ... ABC Action News, WFTS, covers local news in Tampa Bay and Florida.

**Double-parked car blocked with shopping carts - FOX 13 News**
www.fox13news.com/trending/224946987-story
1 day ago - We've all been there. You're in the midst of holiday shopping and the parking lots are full.
Missing: action
You visited this page on 12/22/16.

This brings me to a YouTube video that tells me this was taken "outside a Portland, Oregon Walmart" and has been shared "hundreds of times since yesterday". So back to search. This next result shows you why you always want to look past the first result:
Which one of these items should I click? Again, the idea here is to get “upstream” to something that is closer to the actual event. One way to do that is to find the *earliest* post, and we’ll use that in a future task. But another way to get upstream is to get closer to the event in *space*. Think about it -- who is more likely to get the facts of a local story correct -- the local newspaper or a random blog?

So as I scan the search results I’m looking at the URLs. Fox 13 News has it in “trending”. AmericaNow has it in the “society” section.
But the WGME link has the story in a “news/local/” directory. This is interesting, because the other site said it happened in Oregon, and here the location is clearly Maine. This URL pattern is a strong point in its favor.

Further indications here that it might be a good source is that I see in the blurb it mentions the name of the photographer “Matthew Mills”. The URL plus the specificity of the information tell me this is the way to go.

That links me to what looks like the news page where it went viral, which embeds the or

BIDDEFORD (WGME) -- A photo of a car taking up two parking spots surrounded by shopping carts at the Biddeford Walmart has been shared all over the country.

Matthew Mills of Old Orchard Beach took the picture of the parking lot justice Sunday afternoon.

CBS 13 posted it Monday morning, and it went viral.

That post alone has been shared over 100,000 times and seen by

We see here that the downstream news report we found first had a bunch of things wrong. It wasn’t in Portland, Oregon — it was in Biddeford, which is near Portland, Maine. It hasn’t been shared “hundreds of times” it’s been shared hundreds of thousands of times. And it was made viral by a CBS affiliate, a fact that ABC Action News in Tampa doesn’t mention at all.
OK, let’s go one more step. Let’s look at the Facebook page where Matthew Mills shared it. Part of what we want to see is whether it was viral before CBS picked it up or not. I’d also like to double check that Mills is really from the Biddeford area and see if he was responsible for the shopping carts or just happened upon this scene.

The news post does not link back to the original, so we search on Matthew Mills again, and see some news outlets mentioning the original caption by Mills: “This guy got a lesson in parking”.

That’s not the same as the caption that the news station put up -- maybe it’s what Mills originally used? We pump “got a lesson in parking” Matthew Mills into Facebook, and bingo: we get the original post:
And here’s where we see something I really dislike about news organizations. They cut other news organizations out of the story, every time. So they say this has been shared hundreds of times because in order to say it has been shared hundreds of thousands of times they’d have to mention it was popularized by a CBS affiliate. So they cut CBS out of the story and distort the truth.

On the other hand, one of the good effects of it is sometimes it makes it easier to track something down to the source. News organizations work extra hard to find the original source if it means they can cut other news organizations out of the picture.

But it also tends to distort how virality happens. The picture here did not magically become viral — it became viral due, largely, to the reach of WGME.

Incidentally, we also find answers to other questions in the Matthew Mills version: he took the picture but didn’t arrange the carts, and he really is from Old Orchard Beach.

Just because we’re extra suspicious, we throw the image into Google Image to see if maybe this is a recycled image. Sometimes people take old images and pretend they are theirs -- they just change the supposed date and location. A quick use of Google Reverse image search shows that It does not appear to be the case here, although in doing that we find out this is a very common type of viral photo called a “parking revenge” photo. The specific technique of circling carts around a double-parked car dates back to at least 2012:
When we click through we can see that the practice was popularized, at least to some extent, by Reddit users. See for instance this post from December 2012:

So that’s it. It’s part of a parking revenge meme that dates back at least four years, and popularized by Reddit. It was shot by Matthew Mills in Biddeford, Maine, who was not the one who circled the carts. And it became viral through the re-share provided by a local Maine TV station.

Using Google Reverse Search from Chrome

Most of the time finding the origin of an image on Twitter is easy. Just follow the links. For instance, take the chart in this tweet from Twitter user @NinjaEconomics. Should you evaluate it by figuring out who @NinjaEconomic is?
Nope. Just follow that link to the source. It’s usually the last part of a tweet.

If you do follow that link, the chart is there, with a bunch more information about the data behind it and how it was produced. It’s from the Atlanta Fef, and it’s the Fed -- not @NinjaEconomics -- that you want to evaluate.

But sometimes people will post a photo that has no source, as this person does here:
So I have questions.

First, is this actually a National Geographic photographer?

More importantly, is this real? Is that lava so hot that it will literally set a metal tripod on fire? That seems weird, but I'm no lava expert.

There's no link here, so we're going to use reverse image search. If you're using Google Chrome as a browser (which you should be for this class) put the cursor over the photo and right-click (control-click on a Mac). A “context menu” will pop up and one of the options will be “Search Google for image”.
Doing that you get a bunch of sites that have the photo on them. One of them is Reddit. Reddit is a site that's famous for sharing these sorts of photos, but it also has a reputation for having a user base that is very good at spotting fake photos.
When we go to the Reddit page we find there is an argument over whether the photo is fake or not. But again, Reddit is not our source here -- we need to go further upstream. So we click the link that says it's real and get [taken to article](https://www.reddit.com/r/pics/comments/.../in_the_heat_of_the_moment/) where they actually talk to the photographer:
Now we could stop here, and just read the headline. But all good fact-checkers know that headlines lie. And not just lie. They lie all the time. So we read the article down to the bottom:

For this particular shot, Singson says, “Always trying to be creative, I thought it would be pretty cool (hot!) to take a lava pic with my shoes and tripod on fire while photographing lava.”

This may be a bit pedantic -- but I still don’t know if this was staged. Contrary to the headline the photographer doesn’t say lava made his shoes catch on fire. He says he wanted to take a picture of himself with his shoes on fire while standing on lava.

So did his shoes catch on fire, or did he set them on fire? I do notice at the bottom of this page though that this is just a retelling of an article published elsewhere -- it’s not them who talked to the photographer at all!
In webspeak, “via” means you learned of a story or photo from someone else (our “reporting on reporting” model from The Blaze above). In other words, we still haven’t gotten to the source yet. So we lumber upstream once again, to the PetaPixel site from whence this came. When we go upstream to that site, we find an addendum on the original article:

So a local news outfit has confirmed the photographer did use an accelerant. The photograph was staged. Are we done now?

Not quite. You know what the next step is, right?

Go upstream to Hawaii News Now!

So we do that, we click the link, and we find the quote is good. And I like Hawaii News Now for another reason -- they are a local news service, and so they know a bit about lava fields. That’s probably why they asked the question no one else seemed to, which is essentially “Is that really possible?”

One last check, to check the Hawaii News Now source -- let’s check who Hawaii News Now is. We start by selecting Hawaii News Now and using our Google search option:
And what we get back is pretty promising: there’s a Google Card that comes up that tells us it’s bona fide local news program from a CBS affiliate in Hawaii.

And honestly, you could stop there. We’ve solved this riddle. The photographer was really on hot lava, which is impressive in itself, but used some accelerant (such as lighter fluid) to set his
shoes and tripod on fire. Additionally, the photo was a stunt, and not part of any naturally occurring National Geographic shoot. We’ve traced the story back to its source, found the answer, and got confirmation on the authoritative nature of the source.

We’re sticklers for making absolutely sure of this, so we’re going to go upstream one more time, and click on the Wikipedia link to the article on the Google card to make sure we aren’t missing anything, but we don’t have to make you watch that. I’ll tell you right now it will turn out fine. In this case at least.

Activity: Trace Viral Photos Upstream

These three photos have been attributed to National Geographic shoots as well, by the same tweeter. Find out where these photos were taken and whether they were staged or otherwise faked. For bonus points, get the name of the photographers (or videographers) pictured and if the shoot was associated with National Geographic.

We put the photos below, but to do a reverse image search on them you’ll have to go find them at the Hapgood blog, and use your Google Reverse Image Search right-click.