From Rags to Riches to Whatever the First Google Search is...

Phuc Doan

Did you know that if you grow yourself a meter-long silicon crystal, cut it into tiny millimeter-thin wafers, laser carve into them, and then apply a sprinkle of electricity, you can create a supercomputer that can execute several dozen billion instructions per second? It's a revolutionary feat of engineering that can be simply summed up as magic we chant to look at cat videos! It also happens that computers can be used for other things too, particularly in the information age we live in today where we can now share, process, and interpret so much data at such a rapid rate.

For instance, there are various medical databases built with international effort that are publicly available online; however, Britto-Borges and his team of computational biologists found that there was a lack of accessibility and usability in some of these datasets, particularly in RNA sequences crucial for investigating heart-failure. To address this, the medical researchers compiled Magnetique, which they promoted as "the first online application to provide an interactive view of the HF transcriptome at the RNA isoform level." With their web application, not only were they able to draw some conclusive statements about the underlying biology and chemistry of the datasets, but also shared their program with the rest of the world online – for *anyone* to play around with.

So it should come as no surprise that computers now play a critical role in the education scene. Djamas and fellow researchers, for instance, developed digital multimodal resources that taught Indonesian students about Newton's laws of motion – and not only did it just do that – but the students also displayed an increased enjoyment of the material. The students of all different backgrounds and experiences could now learn at their own pace, made possible by the resources' "student-centered" learning model (Djamas et al.). It allowed students who wanted

a better understanding of a particular concept to independently investigate it further – all without impeding the "unique learning styles" of any other student.

And so, it's not exactly shocking to see that technology has fully crept into our everyday lives, as the use of computers as information processing beasts couldn't be ignored. Britto-Borges addressed the issue of the lack of usable information with their Magnetique web application, and Djamas found encouraging reasons to push for the development of new educational resources. With computers across the world keeping the internet running, it has never been any easier to share information.

With this trend continuing, however, it made me wonder: have we reached an inflection point where there's perhaps an *overabundance* of online resources today? And if so, what qualities would a particular resource have that'd make it more appealing than the dozen other ones? These questions have major implications for how the new generation of students – now so accustomed to the availability of online materials – will interpret and gather technical information. Having dug deep into this subject, I have found that technology's role in delivery has now also emphasized many other subtler aspects of a resource's usability. While traits such as the reputability and reliability of the author are still taken into account, we have now also begun to increasingly put pressure on resources that'll furthermore appeal to our human conveniences.

Literature Review

About 822,000,000 results (0.39 seconds)

The idea that there's perhaps an overabundance of technical resources online is really not too far-fetched, because you can see the effects today on any popular educational forum site. Consider Reddit user u/SurfinShibe7669 who uploaded on r/learnpython the dire request:



Figure 1. u/SurfinShibe7669's post on r/learnpython.

– and in response, over a dozen other Redditors suggested their own favorite introductory Python resource! I often see people like u/SurfinShibe7669 so overwhelmed by the amount of starting points that they turn to others to just *be told where to even begin*. This consequence stems from the fact that producing and sharing a resource has become almost trivial to do; no longer do you need a publisher to push bookstores to shelf your *Python for Dummies* book – just make a blog!

When searching "Python tutorials" gets Google to pull up 822 million results in 390 milliseconds, there becomes the issue of "information explosion" as Gruzova describes in *Overcoming Information Barriers during Technical Communication*. As they put it, when there has been such mass production of a particular topic, it may become *increasingly difficult* to process what they're all actually saying in the conversation. Don't get me wrong – I'm grateful for this development where it's now so easy for me to just search up things about any niche topic I'm interested in, but perhaps life would've also been simpler if there were only one "Python for Dummies" shelved in Barnes and Noble rather than an army of catalogs.

This is a bit of a benign issue for a hobbyist learning Python programming, but it's nonetheless something we should start discussing before it becomes quite serious for our generation – the generation that will innovate fields that are dependent upon technical information itself. From engineers simulating bridge designs to chemists synthesizing new

compounds, our understanding of the world is becoming ever more diverse. In fact, Gruzova presents their concept of "information barriers" – with information explosion being one of them – mostly in the context between professional designers and developers, not self-learning hobbyists, yet both groups are very much affected by the same issue of overabundance.

R.T.F.M.

In search of a solution, one of Gruzova's main conclusions was that many of the information barriers could be greatly mitigated through well-written, standardized technical documentation, but what exactly does that look like, and how well does it actually hold up?

Consider the following page from an integrated circuit datasheet by Texas Instruments (yes, the very same company that makes those ubiquitous TI-84 family of calculators). It might not look like much – just gobbledygook if you're not familiar with electronics – but for someone like me pursuing computer engineering, I find this to be a work of art. It's concise, clear, and tells you as much as the manufacturer *can guarantee about the device*.

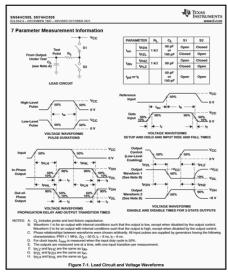


Figure 2. Page 10 of the SNx4HC595 datasheet.

There's literally no other, more authoritative resource that you could refer to than the very one that came from the manufacturer themselves. So not only does this resource avoid Gruzova's information explosion barrier altogether, but it also avoids others like the financial barrier (I didn't have to pay a single penny to access this publicly hosted datasheet!). But even with a peach as sweet as this, it's sometimes not what people *actually* end up using.

For instance, Katriina Byström performed a study with 39 participants from two local Finnish governments with the goal of understanding the relationship between people's subjective perception of the complexity of a task and the types of informational resources they end up utilizing for that task. From the self-recorded journal entries of the municipal officials, Byström collected a total of 54 tasks considered complex ("temporary traffic arrangements to the restoration of damaged landscapes"); the informational resources used for the preparatory work of these matters could be then categorized into three buckets: people, documents, and visits – but since visits were apparently so underutilized – it was just mainly people and documents.

It was found that, as the task demanded more information to be acquired, experts and meetings as human resources became more favorable, in fact, *more favorable* than documentary sources even if the latter was more *available*. Straight from the horse's mouth:

Active information acquisition led directly to a more extensive use of people as sources. This is by no means a novel finding, but **interesting in light of the increased amount of information available in both electronic and traditional documentary sources**[...] This indicates that, in the present setting, documentary sources are not at all as likely to be used as sources [...] as persons are. This is interesting because **a lot of this information is stored in textbooks, manuals, and other guideline documents**.

This is indeed very interesting, Byström! But how come? Well, the study mentioned earlier that "in general, people as sources were useful for the acquisition of all types of information." That is, as I interpret it, people tend to be well-rounded – perhaps not a perfect omnipotent being that knows all the nuanced regulations in Finnish existence – but fully capable of bringing relevant information to the table with very little noise. With documentary sources like guidelines and manuals, however, there's now a potential disconnect between the author's prediction of what is *wanted* and the representative's actual *needs*.

So as much as technology has bridged the cost of delivering information globally, it is of lesser use when there's a discrepancy between the author's expectations and the reader's goals. When both sides manage to be on the same page (no pun intended), then you can have someone like an engineer working confidently with the manufacturer's datasheet; when the "textbooks, manuals, and other guideline documents" (Byström) become too impersonal, we rather fallback on flesh-and-blood people to explain it all to us.

Cognitive Friction

This disconnect in fact is just simply another one of Gruzova's information barriers, where they quote it as a "psychiatric hospital run by its patients." Intrigued by the analogy, I went ahead and got myself a copy of *The Inmates Are Running The Asylum* by Alan Cooper where it originated from. It was from here that the relationship between technology and our ability to gather technical information is actually much more nuanced than originally thought.

Alan Cooper is attributed to being the "Father of Visual Basic", an interactive programming environment for Windows that caught Bill Gates' attention when he demoed it in 1988 ("Why I Am Called 'The Father of Visual Basic'"). From here, he became ever more influential in the design space of software and even later found his own design consultation firm. In his book, Cooper makes the strong claim that computer literacy has become "a euphemism for social and economic apartheid" and that "computer literacy is a key phrase that brutally bifurcates our society" (38).

Perhaps I quite vastly understated when I said "strong", but Cooper isn't really all that far off. Let's tie this back to Djamas developing the multimodal resources to teach Indonesian students about Newton's laws of motion. I brought this up in the introduction to demonstrate the ways in which computers are useful in educational settings, but what I left out was the details of *when it didn't work out!* Specifically, when they field-tested the newly developed interactive multimedia resources, some students didn't have a laptop, and resultantly had to share one with classmates. As the researchers witnessed: "several students felt uncomfortable with this situation, stating they wished to learn with interactive multimedia learning materials using their own laptops and according to their unique learning styles" (Djamas et al.).

Of course, there's a difference between not having a computer and not being computer literate, but both can just be as *handicapping*. Despite computers reducing the cost of information delivery, we are ironically still at the mercy of that technology making the information itself even usable. Without some computer literacy, it can be difficult to articulate where things fall short – and then make up for it – in order to use digital resources effectively.

For instance, I occasionally browse ST Community (the official forum site for my ST-branded microcontroller) to keep up with the current discussion of the latest things. One such thread that I myself have read up on was titled "Hint: DMA and Cache Coherency", where the original poster shared some advice on how to avoid some potentially dangerous issues with memory accesses if certain precautions weren't followed. Threads like these are what make ST Community a rich, informative resource for both engineers and hobbyists like me.

Now, that thread I looked at was posted back in 2016 – and at some point in time – the ST Community forum underwent a "migration" where it was redesigned to be the more modern site you'd see today. Threads before this migration had to be adapted to the new site – and indeed it was done – but quite sloppily! I know this because comments that were too long to transfer were entirely replaced with an error message with a download link to the original text. For the comments that managed to pass through, random parts of it became spuriously italicized or bolded, and horribly formatted code snippets blended in and out with people's sentences. To top it all off, the site would occasionally interrupt your reading by shoving a pop-up in your face asking whether or not you're "satisfied with the ST Community."

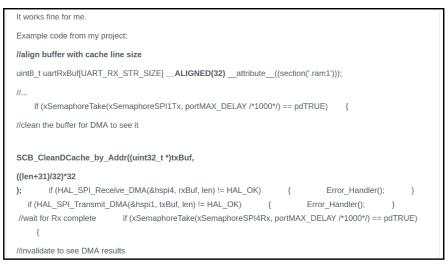


Figure 3. Sample text from a user's comment on ST Community thread.

If I weren't someone who frequents the guts of computers every day, it would've probably otherwise been difficult for me to explain why my experience of using the forum was so subpar. Knowing all of this, however, it's not too hard to understand the meaning of the front cover subtitle of Cooper's book "Why High–Tech Products Drive Us Crazy." While our laptop refuses to connect to

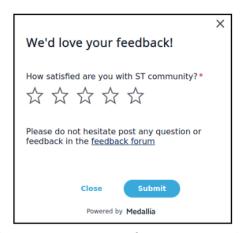


Figure 4. Pop-up survey from ST Community.

the internet and the bloated browser is sucking the battery dry, our interactions with computers are becoming ever more frustrating. Cognitive friction, as Cooper calls it, is the reason why we are now so *overwhelmed* with so many options, buttons, menus, and features of modern software (20).

With software becoming more and more involved in the *content* of informational resources – rather than just delivery – it has now become an additional factor in the quality and effectiveness of the material itself. Sometimes things can work out for the better (like seen with Magnetique as an interactive web application to address the opaque medical datasets) but it can also just as easily be detrimental, as with my experience on the ST Community forum site. Thus,

fundamentally, the way software is crafted has major implications as to what technical resources we end up interacting with and how effective it is in communicating information to us.

We have shed some light on the fact that there's an overwhelming amount of technical resources available today, that literary documentation can be hit-or-miss, and that technology itself plays a huge role – not only just in delivery – but also embedded in the content of the resource itself. I then investigated further into how the new generation keeps their heads above the water in all of this. Specifically, with so many options available, how does the demographic of students decide on what online resource they consult among dozens?

Method

The first primary research I conducted was through a questionnaire that asked students to give their surface-level opinions on various resources. It began by asking basic background questions (year, major, and study habits) followed by three scenarios that involved having to select a particular resource; such a scenario was: "Suppose your professor failed you in explaining how to solve first order linear differential equations, so you decide to consult Professor YouTube. Which of the following videos do you think will be worth your time watching first?"

Since the survey-takers are students of varying backgrounds, expressing the questions in the form of scenarios is intended to help level out the desired intention of the responses. That is, students who were not as well-versed in math, for instance, will choose whatever YouTube video shown in the list to be most appealing – but the ones who were already familiar with first-order linear differential equations would simply choose the video they had already or would've watched when they first started learning the concept. After choosing a particular resource, the survey-taker would then give a short explanation as to why, to which the responses varied from video length to familiarity of the creator, or simply because it was the first thing that was listed.

The goal of this survey is to expose the heuristics in students' minds that evaluate what resources are most appealing to them compared to others. To complement the findings further, I conducted my second research through an interview with a colleague of mine, Lap Le, who's majoring in biochemistry. Being someone who I often always find busy studying, he served as a great representative of industry-leading students who have to confront the potentially overwhelming amount of information in this day and age. The goal of the interview was to find more detail into the exact journey a student goes through to find the technical information they actually need – rather than the surface-level, subjective opinions that the survey would uncover – all in the context of biology, medicine, and chemistry.

Results

Survey

The survey background consisted of 16 undergraduate freshmen and 3 sophomores, with 1 high school junior and 3 seniors. Only 2 out of the 23 were pursuing non-engineering majors with the rest in engineering. 5 survey-takers place studying and being productive as their highest priority, 8 describing it as something they like to do when there's an opportunity, and the remaining 10 are not as concerned.

The first scenario of the survey was the aforementioned example of the YouTube search for "solving first order linear differential equations," to which a list of the first eight results was given. Half of the respondents (13) chose the first video by The Organic Chemistry Tutor, and a quarter (6) chose the fourth video by Blackpenredpen, both popular math content creators. The rest (4) chose one of the remaining six videos in the list of results. Most explanations by the respondents for their choice of resource were categorized into four groups (some cited multiple reasons, to which they are each counted individually). Shy of three quarters, respondents cited familiarity with the content creator, while the remaining considered the video length, placement in the search results, or video thumbnail.

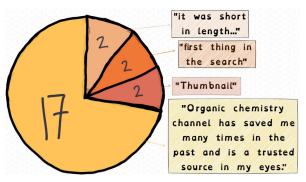


Figure 5. Categorized responses to survey scenario #1. Not to be confused with Pac-Man eating a slice of cheddar.

Following that, survey scenario #2 was: "You decide it's time to finally mature from Excel and start picking up MATLAB programming to create graphs of data you collected from your turbo-encabulator prototype. After a quick Google search, which of the following results would you visit first?" The survey then showed the first six search results, half are official MATLAB sites, and the other half are YouTube videos. About half of the respondents (12) chose one of the three videos while others (7) chose the first MATLAB site.

Most respondents who chose the videos explicitly cite it as being more preferable over textual materials; other reasons such as thumbnail and video length were also mentioned (and are grouped separately). In comparison with the previous scenario, the emphasis on the author of the resource was reduced, while the influence of result rankings increased.

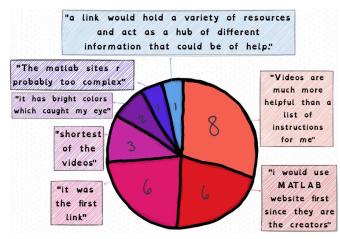


Figure 6. Categorized responses to survey scenario #2.

Lastly, scenario #3's prompt was: "You're an absolute history geek, and this week you want to learn all about the Red Scare of the 20th century, just for funsies! You did some preliminary researching and found some resources. Just by skimming, which of the following will you like to read first?" The resources provided were screenshots of articles by Britannica, UVA, and HISTORY, along with an online PDF of a book and a Reddit thread from r/history, all on the topic of the Red Scare. A plurality of respondents (9) chose Briticanna, while the rest chose evenly across the remaining four sources.

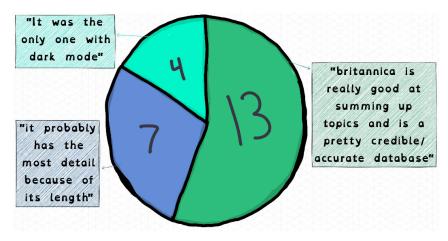


Figure 7. Categorized responses to survey scenario #3.

The majority of respondents prioritized the reputation of the resource first. Others factored it based on the length and amount of details found in the screenshots. Interestingly, several commented on their preferences being solely due to the layout of the resource, usage of pictures, or simply because the site background was dark-themed.

Interview

Lap Le introduced me to his ongoing research of an organic compound called histamine, which can be found in all sorts of biological processes in the human body. He expressed his uncertainties about this particular compound, for instance, the difference between "histamine B" and "histamine E." Thus, he recalled early in his research journey of googling precisely just that: "histamine B vs histamine E" – to which a research paper by the National Institute of Health (NIH) popped up first.

In addition, Lap also introduced me to PyMOL, a program that can visualize and simulate interactions between molecular structures. With this, Lap hoped to recreate some reactions that supposedly unfold with histamines as described in the NIH paper, but as he demonstrated it in front of me: "this is the reaction as it occurs... but they're not reacting with each other..."

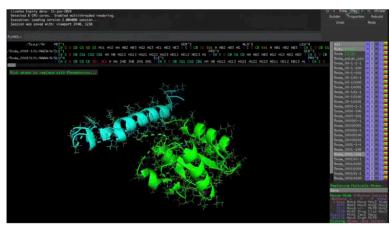


Figure 8. PyMOL simulation between two organic structures.

"I'm trying to find a way to do it," Lap stated his goal. He then showed me the third resource he consulted: the global Protein Data Bank archive (PDB), which stores 3D structure data of various proteins such as DNA, RNA, and – of course – histamine. Lap praised this dataset for being quite reputable, as he said it's internationally curated by those in "biotechnology, biochemistry, biomedical, bio all-kind-of-that!"

Lap then circled back to the research paper provided by NIH to admit that he had some skepticism. His doubt in the paper was in the fact that it was backed by those specifically in biotechnology, while something like the PDB on the other hand is much more multidisciplinary.

"You trust it, but don't as well," not because of nefarious reasons as Lap explains, but that "it can be right, but it also can not be true... because this is like people who are doing research and they write this down, but there's also sometimes when you do it, it doesn't occur like that way." For instance, one of the diagrams in the NIH paper depicted a protein that Lap claimed he recreated in MyPOL – but it also happens to look completely different.

"It doesn't work out..."

"So the way it's simulated was different from what was said in the paper?" I inquired.

"Yeah," Lap defeated.

"What would you trust more, then? The simulation or the paper - or neither?"

After a thought, he declared neither – that he'd actually rather use PDB, for its diverse background of individuals that contribute to it, but also because of its broadness and specificity in its data. Lap showed me the different tabs on the site which each presented tables, graphs, and even renders of the compound – all of which he admits "is pretty cool."

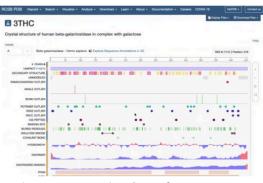


Figure 9. Interactive chart of 3THC on PDB.

In all, Lap summarized his situation succinctly: "YouTube doesn't help with this."

Discussion

The survey shed some light on how certain students evaluate their resources, and it is fortunate for us to say that there is a consistent consideration of reputability across the different scenarios. Outside the theme of ethos, however, the factor of convenience is also quite noticeable; many explanations from respondents cite their choice of resource just being due to the rankings of the search results. Furthermore, only one respondent mentioned the possibility of reaching out to a professor if they needed further guidance. While the survey is not entirely comprehensive, this can indicate for many that the factor of convenience (of watching a YouTube video in the comfort of your own dorm, for instance) heavily outweighs the effort the alternative might otherwise impose (e.g. having to schedule for an in-person meeting).

Additionally, two respondents in the second survey scenario gave somewhat opposing answers. One preferred to use the official MATLAB site to learn, citing that it might "hold a variety of resources and act as a hub of different information that could be of help" while the other stated that the "matlab sites r probably too complex." It's interesting to take note of this as it ties in well with Alan Cooper's idea of cognitive friction and also Byström's usage of subjective complexity. The latter respondent's view that the site is "complex" could be attributed to past experiences of having to deal with similar situations and being frustrated (induced by cognitive friction), and falling back on the preference of a person explaining it (via video format) instead of a documentary text. The former respondent saw past all of that and decidedly might take advantage of the network of resources that technology is able to deliver, which in itself is a skill of computer literacy.

The interview with Lap Le is quite different, for it is a situation where instead of the theme of overabundant resources, there's quite a bit of a shortage. Nonetheless, the three resources Lap utilized (NIH, PyMOL, and PDB) are all quite dense with technical information. Once again, the question of ethos and reputability was quite heavily weighed. Lap's view of NIH as not as reliable as PDB compounding with strange discrepancies with the PyMOL simulations made it difficult for him to shake off his skepticism about all the information he was given. It is interesting to then imagine what would happen if Lap's wish of YouTube videos about his topic came true. Perhaps with many more resources, it'll be easier to have a diverse and comprehensive view of his research – or perhaps the monkey's paw will curl and the whole scene only becomes more confusing as it was for u/SurfinShibe76693.

As stated in the interview, Lap trusted PDB the most, one for its multidisciplinary background, but also for the way it displayed and presented its information to the user. For me, this was quite surprising, as I imagine the PyMOL simulator to be potentially more reliable –

being able to see the entire reaction unfold before you. However, having seen Lap experience difficulties in working the program, I thought of the amount of cognitive friction that is generated. PyMOL and the paper by NIH are by no means made by amateurs, so I speculate that Lap's true confidence in PDB – although never said explicitly – lies in the fact that it was just simply much more user-friendly in providing information.

And so, this appears to be the subtle theme. While we've seen students think critically about the origin of the information, we've also seen decisions made based on human conveniences. This could be picking the first result to spend less energy on having to evaluate other resources, finding a resource to be more appealing because it was easier on the eyes, or the fact that one resource was less complicated to use than the other.

Conclusion

Ideally, we want to believe that we – especially students – are always critically evaluating a resource's recency, relevancy, and credibility to determine if it's worth the attention – but unfortunately, we can't always put that researcher cap on all the time. If there happens to be a sea of options, might as well pick the one with the path of least resistance; there's plenty of fish after all!

It should be noted however that the survey and interview I conducted were only of a small, localized audience, and a more comprehensive sweep should be done to paint a clearer picture. Nonetheless, what we have studied so far could be an interesting starting point for further discussion in other major sources of information such as social media – and perhaps more recently – AI.

Specifically, there's been a noticeable amount of discussion entertaining the idea of artificial intelligences serving as tutors – and after all, we've discussed so far – it's easy to see the appeal! Textbooks, manuals, and articles can be hard to chew through, and the availability

and accessibility to experts can be scarce. AI, however, stands itself as an interesting middle ground between the two; it's been responsive, convenient, and is also impersonal to even the silliest of questions you might ask it. To be able to get such a human-like interaction in discussing a homework problem, revising an essay, *and* explaining any nuanced topic is quite attractive – and even though the technology is still far away from being reliable enough – *it's certainly getting better*.

But I then think back to this one exchange I had with my professor: he talked about how even if we discovered nuclear fusion as an unlimited, clean energy source, the way some of us live our inefficient lives (driving around a humongous SUV in car-dependent infrastructure to order a single latte in the drive-thru of a Starbucks) will still continue. Hell, with such an unlimited energy source, we'd probably be incentivized to mass produce more useless gadgets, manufacture unhealthier foods, and invent earth-splitting bombs. That's all to say, fusion can't guarantee us peace; only we can.

I was thinking about that exchange because of this rise of information technologies – the internet, social media, and now AI – they all have the same ring to it. Our generation has been born and raised into the age where they now have the ability to communicate to such a wide audience in such unique and diverse ways. We live in exciting times for sure, but we must also not forget that these recent advancements are only mere tools; it's by no means an excuse for us to forget to be great communicators of science, of morals, of principles. Fear that we speak our knowledge, ideas, and voice to the whole world – only for it to be the sound of a raindrop in the ocean.

Works Cited

- Britto-Borges, Thiago, et al. "Magnetique: An Interactive Web Application to Explore Transcriptome Signatures of Heart Failure." *Journal of Translational Medicine*, vol. 20, no. 1, 2022, <u>https://doi.org/10.1186/s12967-022-03694-z</u>.
- Byström, Katriina. "Information and Information Sources in Tasks of Varying Complexity." Journal of the American Society for Information Science and Technology, vol. 53, no. 7, 2002, pp. 581–591, https://doi.org/10.1002/asi.10064.
- Cooper, Alan. "The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. 2nd ed." Sams Publishing, 2004.
- Djamas, Djusmaini, and Vonny Tinedi. "Development of Interactive Multimedia Learning Materials for Improving Critical Thinking Skills." *International Journal of Information and Communication Technology Education*, vol. 14, no. 4, Oct. 2018. Gale OneFile: Educator's Reference Complete,

https://link.gale.com/apps/doc/A759563523/PROF?u=viva_vpi&sid=bookmark-PROF& xid=2fc2fc06. Accessed 21 Feb. 2024.

- Gruzova, A. A. "Overcoming Information Barriers during Technical Communication." *Scientific* and Technical Information Processing, vol. 46, no. 2, 2019, pp. 90–99, https://doi.org/10.3103/S0147688219020102.
- "Hint: DMA and Cache Coherency." STMicroelectronics Community, 20 June 2018, community.st.com/t5/stm32-mcus-products/hint-dma-and-cache-coherency/td-p/4 39559. Accessed 24 Feb. 2024.
- "SNx4HC595 8-Bit Shift Registers With 3-State Output Registers." Texas Instrument, Oct. 2021, https://www.ti.com/lit/ds/symlink/sn54hc595-sp.pdf. Accessed 24 Feb. 2024.

SurfinShibe76693. "R/Learnpython on Reddit: What Is the Best Resource to Actually …" Reddit, www.reddit.com/r/learnpython/comments/17zjeig/what_is_the_best_resource_to_a ctually_get_started/. Accessed 23 Feb. 2024.

"Why I Am Called 'The Father of Visual Basic.'" Cooper Software Inc.,

web.archive.org/web/19961101084552/www.cooper.com/alan/father_of_vb.html. Accessed 30 Mar. 2024.