[Mod] Recruiting volunteers - please read

August 5, 2014 | 29 upvotes | by redpillschool

Archived from theredarchive.com

Comments

FrogTrainer • 10 points • 5 August, 2014 04:40 PM

Would it make sense to put a project such as this on Github?

That way you are at less risk of losing work should a dev drop the project before he can finish, and multiple people could submit to it/review it, etc.

I might be able to contribute but I don't want to make any large time commitments at this point. But it would be nice to have a Github repo we could all look at.

We'd probably want to create a extra accounts on GH to prevent doxxing. My normal account is heavily tied to my work projects.

> [deleted] • 4 points • 5 August, 2014 05:31 PM
>
> Git hub it so anyone can contribute. I, myself, might want to contribute a few lines of code, and we could have lots of code auditing to prevent some sort of attack on us.
>
> > RedPillington • 8 points • 5 August, 2014 11:13 PM
> >
> > that's presuming feminists can code
> >
> > > jolly--roger • 13 points • 5 August, 2014 11:15 PM
> > >
> > > They have tons of betas who can

antwonedw • 2 points • 5 August, 2014 05:21 PM

Definitely interested. Experienced sr software engineer here. Replying instead of pm cause on phone.

CokeandGrappa • 2 points • 5 August, 2014 10:46 PM

I am not thrilled by the "offline archival", but i got the qualities you ask for. Care to elaborate your intentions?

> redpillschool[S] • 8 points • 5 August, 2014 10:48 PM
>
> Just looking to back up some of our best posts and discussions in a rare case of emergency.
>
> > erqos • 1 point • 6 August, 2014 12:21 AM
> >
> > > to back up some of our best posts and discussions
> >
> > How does one determine what's "best"? Are you going to go down in descending order by link karma points? Because I've seen really good posts and comments that were neglected in terms of votes for various reasons.
> >
> > Also, I'm not opposed to a coordinated and centralized effort, but it'd probably be a good thing for multiple groups or individuals to spearhead their own distinct archival efforts, based on their own preferences. Also we need a much better query system than what reddit's providing, so as to facilitate specific lookups based on specific criteria. But yeah, regardless of how we do it, if we lose this treasure trove of knowledge, it'd be equivalent to when Alexandria was burned down.
> >
> > > [deleted] • 2 points • 6 August, 2014 03:18 AM
> > >
> > > Why don't we just put it in a torrent?

trpill • 1 point • 6 August, 2014 09:12 AM

You can back this entire sub easily. There are people who back all of GW.

Venture to /r/datahoarder for more information

QuietlyLearning • 2 points • 5 August, 2014 11:38 PM

Would we be able to download said archive using something like a torrent since I presume it would be large?

AFPJ • 2 points • 6 August, 2014 03:50 AM

Every post and comment ever made to this sub fit into less than 3GB of UTF-8.

jolly--roger • 1 point • 6 August, 2014 07:32 AM

3GB?

That's a fucking ludicrously huge amount of text. A few libraries perhaps.

We're in the range of MBs at most, and text is easily compressed as well.

AFPJ • 1 point • 6 August, 2014 10:35 AM*

3GB is only 3,221,225,472 characters. Now factor in indexing, XML/HTML encapsulation & XDATA overhead. The "Men Are Not Happy" post is among the largest on this sub and its source contains 809,335 Characters circa this comment - only 3,980 of those would fit into 3GB. A lot of space is saved via compression & redacting superfluous / irrelevant HTML into XML but 3GB is a miniscule amount of data, even for plaintext.

You'd have to redact the text to a fucked up encoding like ASCII, use draconian delimiters such as CDATA or straight up line breaks and then compress everything with a SOTA algorithm like BWT, PPM or Dynamic Markov in order to preserve this sub in "the range of MBs". Why would anyone ever want to write software to do this when a few gigs costs a few pennies? **OP isn't trying to transmit TRP to the mars rover.**

jolly--roger • 1 point • 6 August, 2014 12:23 PM

> only 3,980 of those would fit into 3GB

And that's about 3700 articles full of noise, actual reposts or semantical reposts, covering say hypergamy over and over (and over again), like we didn't catch it the first time.

Please, don't try to tell me you'd actually need 3GB to save the essence of TRP. You would not.

I think it's pointless to backup the whole sub. I'm not saying it's not easy, it's just pointless.

AFPJ • 1 point • 6 August, 2014 01:39 PM

> I think it's pointless to backup the whole sub. I'm not saying it's not easy, it's just pointless.

Thread removed so looks like OP found whatever he was looking for, but it's easier to archive the whole sub than run PageRank or some other kind of Fuzzy Logic / AI or even worse, manually hand-pick posts.

jolly--roger • 1 point • 6 August, 2014 01:58 PM

it's not removed, it's no longer stickied.

thanks for regurgitating what I already said -- ie. making your comment and this reply

extremely pointless.

go ahead and pick quantity whenever it suits you. I pick quality.

> AFPJ • 1 point • 6 August, 2014 02:04 PM
>
> I know it's 6AM and you probably haven't had coffee, but at least pretend to pay attention.
>
> You said
>
> > it's just pointless.
>
> I said
>
> > but it's easier
>
> Because your logic of
>
> > pick quantity whenever it suits you. I pick quality.
>
> Is faulty. You get all the quantity without losing any of the quality with backups.
>
> It's an automated process that doesn't require human intervention or hand-picking.
>
> > jolly--roger • 1 point • 6 August, 2014 02:06 PM
> >
> > It's 4 pm here.
> >
> > You make assumptions and, apparently, wrong ones.
> >
> > > AFPJ • 1 point • 6 August, 2014 02:15 PM
> > >
> > > Bro, just stop it hahah. http://en.wikipedia.org/wiki/Straw_man
> > >
> > > Why are you still replying?

FrogTrainer • 0 points • 6 August, 2014 07:47 PM

Where did you pull the 3GB stat from?

[deleted] • 1 point • 6 August, 2014 12:46 PM

TheRedFile, I like the ring of that.

[deleted] • 1 point • 5 August, 2014 04:28 PM

If you are doing an archival service, could you please make a Syndie archive? Such as if we're blocked in the UK as a hate site and so people can access our material through the darknet. Essentially, it allows you to distribute your backup to multiple people who then can synchronize it with their own copies, which they can distribute ad nauseam. But then you can post and it'll go to everything else. Unfortunately, it's styled after regular forums. https://syndie.i2p2.de/