

Mozilla Common Voice

BY ERIBERTO GUZMAN, IFTIKHAR KHAN, AND SUZEN FYLKE
FOR CSCI 39586 - OPEN SOURCE SOFTWARE DEVELOPMENT
PRESENTED ON 14 MAY 2019

LICENSE: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/))

What is Mozilla Common Voice?



Common Voice is Mozilla's initiative to help teach machines how real people speak.

- Its goal is to make voice recognition **open** and **accessible** to everyone and to promote **innovation** of voice systems in doing so.
- To date, it is **the largest public domain transcribed voice dataset**. It includes **18 languages** with voice clips from **more than 42,000 people**.
- People can get involved by **reading and validating sentences** and by **improving the website and the tech that powers it**.

(Learn more at the [Common Voice website](#) and the [Mozilla blog](#).)

Why Mozilla Common Voice?



We originally chose Nextcloud Server, but it required


- Having a specific operating system or using a VM,
- Working with a very large codebase,
- Knowing how to troubleshoot server errors, and
- Being comfortable using IRC.

So we switched to Mozilla Common Voice!

Why Mozilla Common Voice?



**We believe in Mozilla's mission,
the project aligns with our
personal interests, and the
codebase was smaller and
seemed more approachable.**



Issues and Fixes

Our Approach

Since we each had to submit a pull request, we brainstormed issues together and had each person take the lead on an issue.

The Issues



We chose issues tagged "help wanted". These are approved by maintainers and are "up for grabs".

- Validate more than 5 recordings (#1869) - Eriberto Guzman
- Datasets do not download in Chrome (#2025) - Iftikhar Khan
- Localize Open Graph description tag (#1051) - Suzen Fylke

**Validate more than 5
recordings**

(#1869)

Problem

By default, users can only validate 5 recordings at a time. For users who have put aside time to validate a large number of recordings, clicking "Ready to do 5 more" is a nuisance and waste of time.

Possible Approaches

- If you press p for play, "Ready to do 5 more" is pressed.
- Users can remove the summary screen in their profile settings to be presented with another five clips immediately.
- **Eriberto noticed that if you press enter, "Ready to do 5 more" is pressed.**

Result

Eriberto suggested instructing users to press enter at the summary screen, and to link to the profile setting that automatically bypasses the summary screen from the corner of this page, where it would be more visible.

He wrote this on the issue, weeks ago, but no one has responded since.

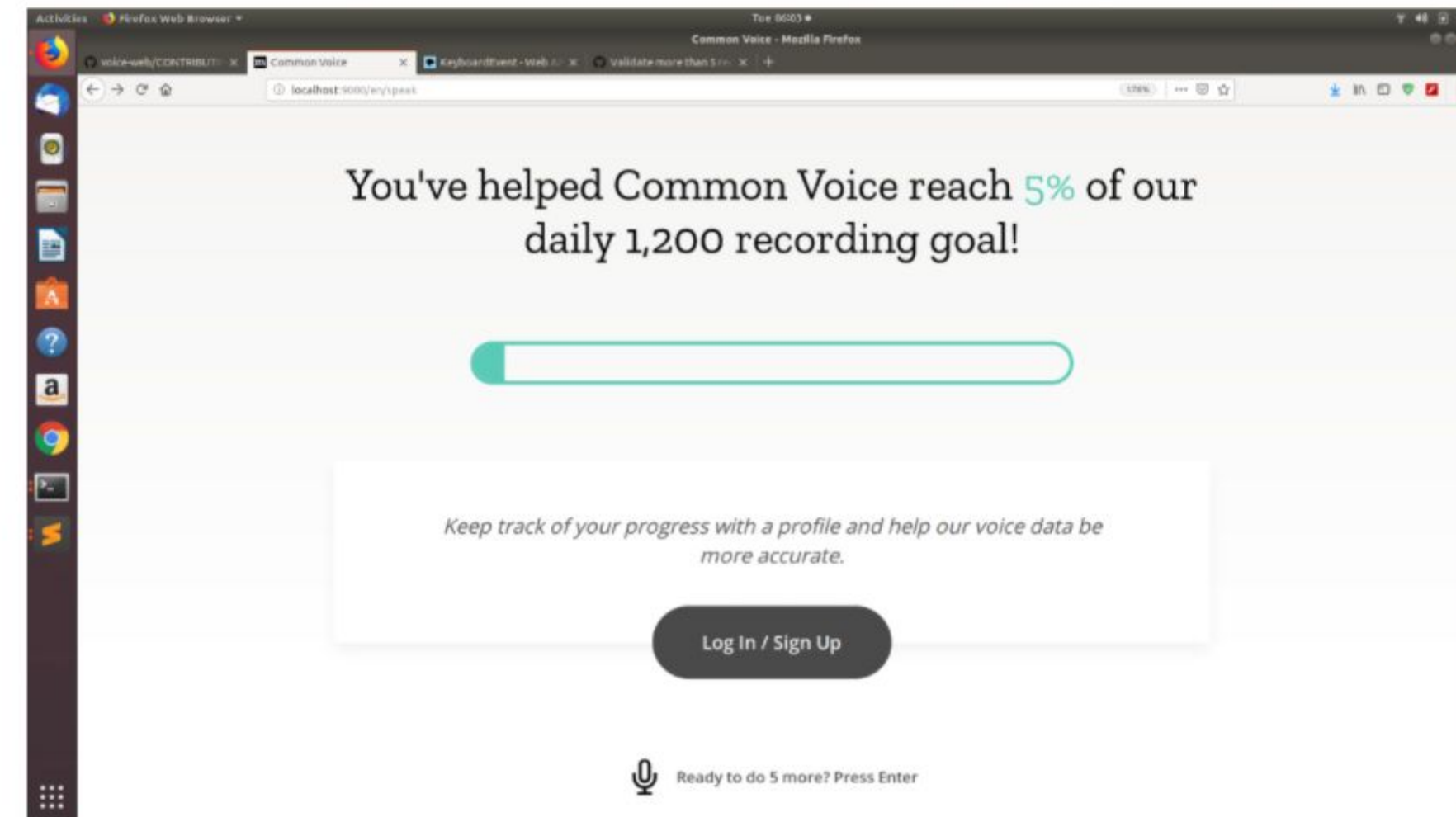


Ejguzm19 commented 13 days ago



Hi,

I was looking at this issue and noticed that if you press Enter it automatically does 5 more. Instead of coding it to pressing P why not just add "Press Enter" to the button.



As for the option to remove the summary screen in your profile settings, I was trying to code the button on the settings page to the actual summary page but was unsuccessful. I think putting it in the corner would improve the visibility.

Datasets not downloading in Chrome

([#2025](#))

Problem

The checkboxes do not show in Chrome that are required to download the datasets.

The issue only appeared in Chrome and reproduced it in Safari, confirming it to be a WebKit issue

Possible Approaches

- Update application State when checkbox is clicked
- SVG rendering bypass
- Set WebKit to not render the element with domain-specific CSS
- Find when issue was introduced

Result

The state not updating was a false alarm because of how Node.js works.

The SVG <path> rendered in a standalone app for both browsers.

Domain-specific CSS did not change anything.

The issue was fixed by changing the color of the custom checkmark from pure black to “near black”

New issue



Collaborator +  ...

This only happens in Chrome, not in Firefox.

Assignees

No one assigned

Labels

Type: Bug

Projects

None yet

Milestone

No milestone

Notifications

🔊x Unsubscribe

You're receiving notifications because you commented.

2 participants



Look to this page as a reference hub for other open source voice datasets and, as Common Voice continues to grow, a home for our release updates.



LRSaunders4 added the

Type: Bug

label 5 days ago



f0cus10 commented 3 days ago • edited ▾

Contributor



Can confirm this issue exists in Chromium 74.0.3729.131 in **both** the staging and release server.

I could not reproduce the issue in Firefox 68.0a1 Nightly .

~~The attributes are being passed as read-only for the checkbox elements.~~

EDIT: After initial inspection, it seems that the error occurs because chrome cannot render the CheckIcon for some reason.

Accidental Bisect

```
ui — -bash — 99x24
icons      icons.tsx      ui.css      ui.tsx
[Iftikhars-MBP:ui focus10$ clear

[Iftikhars-MBP:ui focus10$ git blame -L 165,+10 ui.css
cf93a8e80 (Gregor Weber 2019-02-15 16:45:00 +0100 165)      display: none;
cf93a8e80 (Gregor Weber 2019-02-15 16:45:00 +0100 166)      position: absolute;
69ac13bed (Gregor Weber 2019-04-24 10:36:08 +0200 167)      top: -6px;
5bb51d688 (L#          2019-02-25 16:04:14 +0100 168)
5bb51d688 (L#          2019-02-25 16:04:14 +0100 169)      & path {
3271508c0 (Gregor Weber 2019-04-29 15:08:49 +0200 170)      fill: var(--black);
5bb51d688 (L#          2019-02-25 16:04:14 +0100 171)      }
cf93a8e80 (Gregor Weber 2019-02-15 16:45:00 +0100 172)      }
cf93a8e80 (Gregor Weber 2019-02-15 16:45:00 +0100 173)
cf93a8e80 (Gregor Weber 2019-02-15 16:45:00 +0100 174)      & input:checked ~ .checkmark {
Iftikhars-MBP:ui focus10$
```


Merged Pull Request ([#2033](#))

Revert fill to near-black for visibility in Chrome & Safari #2033

 **Merged** Gregoor merged 1 commit into `mozilla:master` from `f0cus10:checkbox-fill` 20 hours ago

 Conversation 0  Commits 1  Checks 0  Files changed 1



f0cus10 commented 2 days ago

Contributor +  ...



- Closes [#2025](#)

The color had to be changed from `--black` to `--near-black` because:

1. Both `Safari` & `Chrome` fail to show the **checkmark**.
2. Using `-webkit-fill: var(--near-black)` to create a specialized case for *WebKit* browsers and retaining style for others did **NOT** work.

This essentially undoes [3271508](#) .


  Revert fill to near-black for visibility in chrome ✓ 64e650f

  Gregoor merged commit **cd9115a** into `mozilla:master` 20 hours ago
7 checks passed

[View details](#)

[Revert](#)

Path to Pull Request



viorelaioia commented 14 days ago

Collaborator + 🗨️ ⋮


Env:
iPhone 7, iOS 12.2, Firefox and Safari

STR:


1. Login to CV on mobile
2. Navigate to Stats page
3. Tap "Create a custom goal" on Speak section
4. Scroll down and tap "Set a goal" button

Expected:
Next page is shown from the beginning

Actual:
Next page is scrolled down, as the previous one. User needs to scroll up to see the first part of the page and to actually set the type of the personal goal. See [video](#)




viorelaioia added the **Type: Bug** label 14 days ago



f0cus10 commented a day ago

Contributor + 🗨️ ⋮


Is development even active on the iOS App? Last commits are from a year ago



Gregoor commented 21 hours ago

Collaborator + 🗨️ ⋮

Yes-ish, the iOS app is just a wrapper for the webview, which uses native APIs for recording (as Safari has not implemented them YET). So it gets "updated" anytime we update the site itself.




f0cus10 commented on Mar 31 • edited ▾

Contributor + 🗨️ ⋮

Not entirely sure, but maybe archive.org? They create a torrent too while supplying free downloads.

@missuniverse I did some digging and turns out archive.org items cannot have more than **100,000** files. It would have been great if we could since it provides a command line tool that we could use to automate the flow.


Edit: there are 677,021 `mp3` files in the `en` collection



missuniverse commented on Mar 31

Contributor + 🗨️ ⋮


@f0cus10 Then its just a matter of how we package it. If the files are in an archive(tar,zip,rar...etc), then we are good. Also, to my knowledge archive.org partners with organizations to archive their stuff which I think might be cheaper then s3. They helped NASA JPL to archive their public images.



f0cus10 commented on Mar 31

Contributor + 🗨️ ⋮

Yeah, I was thinking that might be a way to bypass those limitations. I'll keep on digging in that area.




Gregoor commented on Sep 20, 2018 • edited by open-innovation-unito-bot ▾

Collaborator + 🗨️ ⋮


The biggest hurdle I see towards contributing to the code is the amount of work required to setup the backend. Docker should've solved that for us, but it has its own problems (one of which is that I don't use it myself for development). Also the largest share of our issues are frontend related. So I think it'd be valuable to look into a frontend-only setup that would either talk to a backend mock or our staging server.

Steps necessary:

- CORS on staging
- proxy API requests to staging
- split up the FE & BE build / dev tasks
- document it



Gregoor added the **help wanted** label on Sep 20, 2018




f0cus10 commented on Mar 31 • edited ▾

Contributor + 🗨️ ⋮

I think this is something interesting.
If I understand the issue correctly, you're suggesting

1. Setting up a CORS configuration on S3
2. Proxying the front-end so it communicates with S3 staging server instead of searching for the backend
3. And make a build script for the client side only



Gregoor commented 29 days ago


Author Collaborator + 🗨️ ⋮

Yes, that sounds about right :)

add UI feedback when using yes/no keyboard shortcuts

#1257


Open Djfe opened this issue on Jul 12, 2018 · 3 comments



Djfe commented on Jul 12, 2018 • edited by open-innovation-unito-bot ▾

Contributor + 🗨️ ⋮


shortcuts shouldn't only blink green/red when being clicked by cursors but also when shortcuts are used.
the easiest solution might be to use the DOM and execute the click method on a button, which in return should also trigger the css, right?



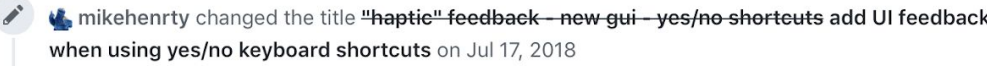
Gregoor commented on Jul 13, 2018

Collaborator + 🗨️ ⋮


Thanks for raising this, that feedback would indeed be valuable. Unfortunately it wouldn't be as simple as simulating a click as as that style is currently only applied on hover.



mikehenrty added the **help wanted** label on Jul 17, 2018



mikehenrty changed the title "~~haptic feedback - new gui - yes/no shortcuts~~" to **add UI feedback when using yes/no keyboard shortcuts** on Jul 17, 2018



Undigon commented on Jul 19, 2018

Contributor + 🗨️ ⋮

Well, we can try using `<input type="checkbox">` and a bit of CSS+JS to implement onClic styling. We should then use some method to remove the checked property once the phrase has been updated. Is there any listener for that or should we apply a timed delay?

I've tried doing something like that before so I know it could be done (changing the colour "on checked" at least).

For reference, I got the idea some months ago from here: [Codepen example](#).

Localize Open Graph description tag

([#1051](#))

Problem

Mozilla Common Voice is a global platform, so it's important to localize content for users. Currently, when someone shares a link to the site with any locale, the description is in English rather than the language of the specified locale. For example, if someone posts a link to the French version of the site on Facebook, the description of the site will be in English, and it might not be accessible to the intended audience.

Possible Approaches

- Get the official translation of the site description in the language of each locale.
- Add the site description string to each locale.
- Add the og:locale:alternate tag to the index.
- Set the content attribute of the open graph description according to the specified locale.

Result

Suzen made a GitHub Gist of her proposed changes and shared them with the Common Voice Slack, but no one responded for weeks, despite the fact that she pinged them multiple times.

In the end, she was told that though this was the correct approach, it would not be possible with the current project structure, in which there's one shared index for every language.

Path to a Solution

Add **og:locale:alternate** tag to the index.

```
<meta property="og:locale" content="en" />
<meta property="og:locale:alternate" content="sv-SE" />
```

Add **site-description** string to each locale's strings.

```
## Metadata
site-description = This is the Swedish version of the site description.
```

```
## Metadata
site-description = This is the English version of the site description.
```

Update **og:locale** tag and **og:description** tag in the index (via the script that builds it) to force Open Graph to reparse the page.

```
document.querySelector('meta[property="og:locale"]').setAttribute('content', mainLocale);
document.querySelector('meta[name="description"]').setAttribute('content', 'Hello!');
document.querySelector('meta[property="og:description"]').setAttribute('content', 'Hello, OG!');
```

This is the necessary approach when using the Open Graph API, but it won't work because the same index is used for each locale. Users are presented the correct locale by default. They never have to fall back to the alternate, Open Graph isn't forced to reparse the page, but this is the only way that Open Graph respects client side changes to metadata.

Threads



general

gregor and you



Suzen Fylke 1 month ago

I'm working on issue 1051 (localizing the open graph description tag), and I've gotten stuck on two problems:

1. When I update the tags with JavaScript, these changes are reflected in the DOM, but they're not picked up by the open graph parser. I'm using the Open Graph Preview Chrome extension.

2. I don't understand how to supply a message from a messages.ftl file to index_template.html.

Here's a gist of what I've done so far: <https://gist.github.com/codesue/85e6bd6f4fb35d06e60c6e6d09bbe788>

Am I going in the right direction with this?

#+ Also sent to the channel



Suzen Fylke 1 month ago

(Just wanted to bump this. 😊)



1



Suzen Fylke 1 month ago

@nukeador

#+ Also sent to the channel



gregor 13 days ago

Oh sorry, I completely missed this. Feel free to @-me directly in the future.

Yeah I think it wouldn't work through JS, we'd need to bring it into the index file like you said. Which might not be possible with the way it's currently built as we have one file for every language



Suzen Fylke 11 days ago

Thank you!




Reply...



Successful Contributions to Other Projects

Merged GSA/code-gov-style: Add a pull request template ([#73](#))

Add a pull request template #73


 **Merged** saracope merged 1 commit into [GSA:master](#) from [codesue:add-pr-template](#) 2 hours ago

Conversation 1

Commits 1

Checks 0

Files changed 1



codesue commented a day ago

Contributor + 😊 ...

Summary
Added pull_request_template.md in the .github folder using the template on code-gov-front-end as a guide.

This PR fixes/implements the following **bugs/features**

- ☒ A pull request template

Explain the **motivation** for making this change. What existing problem does the pull request solve? This change makes it easier for contributors to communicate relevant information, and it saves time because the template auto populates.

Test plan (required)
Visual: navigate to "Pull requests" tab and click "New pull request" button.

Code formatting

Closing issues

Closes [#72](#)

Successful Contributions to Other Projects

Merged WordPress/gutenberg: Change block-level @link tag to @see tag ([#15616](#))

Fix: change block-level @link tag to @see tag #15616

Merged nosolosw merged 1 commit into WordPress:master from codesue:fix/correct-link-tag-usage 40

Conversation 3Commits 1Checks 1Files changed 10

codesue commented 14 hours agoContributor

Description

- Replaced the @link tag with the @see tag when @link was used as a block level tag.
- Did not change the tests and data in `get-jsdoc-from-token.js` or in `fixtures/tags-function/exports.json`.

This fixes part of issue [#14334](#).

How has this been tested?

Visually

Screenshots

Types of changes

Bug fix (non-breaking change which fixes an issue)

Checklist:

- ☒ My code is tested.
- ☒ My code follows the WordPress code style.
- ☒ My code follows the accessibility standards.
- ☒ My code has proper inline documentation.
- ☒ I've included developer documentation if appropriate.

👍 1

Accomplishments

Challenges

Little communication and interaction

Delayed response time

Insufficient guidance

Lacking documentation

Inadequate issue labeling and organization

Non-public hosting

Difficult environment setup for some

Proposed Improvements

Questions?

Thank you!