

Taking an Entire Page Offline using the HTML5 FileSystem API



By Eric Bidelman

Engineer @ Google working on web tooling: Headless Chrome, Puppeteer, Lighthouse

Let's face it, AppCache is annoying and has problems [1, 2, 3]. One big limitation is the fact that it's impossible to dynamically cache assets on demand. Essentially, this makes it an all or nothing thing when it comes to taking an app offline. Either everything in the manifest is cached up front, or nothing is cached.

The HTML5 FileSystem API [\[4\]](#) becomes an attractive solution for AppCache's shortcomings. One can programmatically store files and folder hierarchies in the local (sandboxed) filesystem and subsequently add/update/remove individual resources as necessary. My colleague, Boris Smus, even wrote a nice library to manage this type of offline caching in the context of games. The same idea can be extrapolated to work with any type of web app.

crbug.com/89271 is an important fix for the FileSystem API which makes relative `filesystem: URL` paths work like a charm.

Let's say for example, I've saved `index.html` in the filesystem's root folder (`fs.root`), created an `img` folder, and saved `"test.png"` in it. The `filesystem: URL` for those two files would be `filesystem:http://example.com/temporary/index.html` and `filesystem:http://example.com/temporary/img/test.png`, respectively. Then, if I wanted to use `"test.png"` for an `img.src`, I needed to use its full absolute path: ``. That meant rewriting all of the relative urls in `index.html` to point to corresponding file's `filesystem: URL`. Not cool! Now, with this bug fix, I can keep the relative path to the file (``) as they will resolve correctly to a filesystem origin.

This feature makes it trivial to pull down a page and save all of its resources offline, while still preserving the exact same folder structure as the online version.

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 3.0 License](#), and code samples are licensed under the [Apache 2.0 License](#). For details, see our [Site Policies](#). Java is a registered trademark of Oracle and/or its affiliates.

Last updated July 2, 2018.