INTERNETARCHIVEBOT

ABOUT THE BOT AND WHAT IT TAKES TO SET IT UP

A LITTLE HISTORY

- The project was first started in May 2015
  - I was pretty bored and was looking around the English Wikipedia bot requests pages.
  - Back then it ran under Cyberbot II
- The bot's code was just a little over 500 lines of code.
- Only designed to run on enwiki
- Could not detect dead links, only links marked dead.

A LITTLE HISTORY

- It made top ten in the community tech wish list.

- On January 30, 2016 the dead link checking algorithm began development with assistance from the Community Tech Team of the Wikimedia Foundation. Shout out to Niharika, Kaldari, and MusikAnimal.

- On July 13, 2016, the bot code was moved to the InternetArchiveBot account.

- But it turns out scanning for dead links is not as trivial as simply checking for a status code.

A LITTLE HISTORY

- The bot start growing more complex as we tried to have it better handle the numerous ways references are formatted on wiki.

- In February 2017, I officially started working for Internet Archive to further develop InternetArchiveBot

- In April 2017, the bot was launched on the Swedish Wikipedia as part of IABot v1.3. The aim is to have the bot run on every wiki.

- It was also the debut of the new UI for IABot found out https://iabot.toolforge.org

A LITTLE HISTORY

- This presented a problem as the bot's original code was only built specifically for enwiki and required code modifications for every wiki.

- Scalability became an issue, as did resource efficiency. The bot still needed a lot of RAM and CPU to actually do it's job.

A LITTLE HISTORY

- By IABot v1.4 – nlwiki, nowiki, zhwiki, and Wikispecies were added to the bot's supported wiki, each with code additions.

- By IABot v1.5 – alswiki, ckbwiki, dewiki, eswiki, frwiki, itwiki, jawiki, and ruwiki were added.

- v1.6 did not add any additional wikis

A LITTLE HISTORY

- The code was starting to become difficult to maintain in its current state.

- IABot v2 seeked to address this by removing wiki specific customizations in the code and migrating them to a configurable engine. Each wiki could run the same could on different configuration.

- And it worked…. Mostly.

- The new code is also much more efficient in system resources.

SO WHERE ARE WE NOW?

- The bot is setup and running on more than 100+ wikis now.

- The bot in most cases doesn't need specialty code to run on a certain language Wikipedia.

- We have basic support for Wikidata.

- Some wikis are even proactively archiving instead of reactively
  - Chinese Wikipedia
  - Russian Wikipedia
  - Turkish Wikipedia
  - Ukrainian Wikipedia

SO WHERE ARE WE NOW?

- As of this morning, IABot has made17,027,661 edits to articles on 100+ wikis.

- Of those page edits, 8,636,531 were made in an effort to fix a broken link.

- IABot has fixed about 12,842,902 broken links.

- IABot has added archive URLs to about 22,169,900 links.

- All of these numbers are not unique.

WHAT IS INTERNETARCHIVEBOT?

- In it's simplest form it's a bot that tries to fix broken links.

- It is a bot that can operate on multiple wikis, on multiple servers, all on one installation.

- It offers a UI for end-users and operators to allow for configuring the bot, and even running it.

- It can intelligently learn about Citation templates, and can integrate with the widely used CS1 Module.

WHAT IS INTERNETARCHIVEBOT?

- It supports many archives:
  - Internet Archive's Wayback Machine
  - Europarchive
  - Archive.today
  - Memento
  - Web Citation
  - York University
  - Archive-It
  - Arquivo.pt
  - Library of Congress
  - The National Archives
  - Bibliotheca Alexandrina
  - Library and Archives Canada
  - Eesti Veebiarhiv

- Vefsafn.is
- The Public Record Office of Northern Ireland
- Univerza v Ljubljani
- Stanford University
- UK Parliament
- National Library Board Singapore
- Perma.CC
- Google Webcache (please don't use this)
- National Library of Australia
- Wikiwix
- Freeze Page
- Webrecorder
- UK Web Archive
- Ghost Archive

HOW TO USE IT (DEMO)

SETTING IT UP FROM SCRATCH

- For users here that would like to run the bot on a privately installed wiki, the bot can be installed in almost any environment.
    - Some technical experience is needed to get it set up.
- Set up requirements:
    - PHP 7.4 or higher with the following dependencies php-mysql, php-mysqlnd, php-json, php-intl, php-pcntl, php-curl, php-mbstring, php-simplexml, php-tideways, php-xml, and tor
    - Composer
    - A webserver, like Apache or Nginx
    - MySQL or MariaDB

QUICK EXPLANATION DEMO OF NEW INSTALL

- Not going to get into too many specifics about installing a new copy of IABot onto an environment.

- The assumption here is if you can operate a MediaWiki installation, you can probably install IABot without a step by step guide.

- For most Wikimedians, this is not necessary.

WHAT DO WIKIMEDIANS NEED TO DO?

- Fortunately, not a lot. As the operators we will do the technical leg work of getting the bot set up.

- All that's needed is the user to help out with some localization work

- Point us to bot approvals, if needed and help us get a bot approval request started.

- James Hare will demo what this looks like.