Page   Discussion   Read   View source   View history   Search Wikitech   🔍

Toolforge webservices are in the final stages of migrating to the toolforge.org domain .
Please help us clean up older documentation referring to tools.wmflabs.org!

Incident documentation/20170512-API

< Incident documentation

Contents [hide]

1 Summary
2 Timeline
3 Other service impacts because of the core change
4 Conclusions
5 Actionables

Summary

The TextExtracts extension (which provides the `extracts` prop to the MediaWiki API) was providing empty `extract` fields for queries for a period of 18 hours. This affected RESTBase and subsequently mobileapps.

Timeline

- **14:26:30**: Icinga begins reporting problems with RESTBase

  ```
  PROBLEM - restbase endpoints health on restbase1007 is CRITICAL:
  /en.wikipedia.org/v1/page/random/{format} (Random title redirect) is
  CRITICAL: Test Random title redirect returned the unexpected status 503
  (expecting: 303): /en.wikipedia.org/v1/page/summary/{title} (Get summary from
  storage) is WARNING: Test Get summary from storage responds with unexpected
  body: /extract = : /en.wikipedia.org/v1/page/title/{title} (Get rev by title
  ```

- **14:27:34**: `bblack` , `_Joe_` , `godog` , `elukey` , and `mobrovac` begin investigation in IRC
- **15:06:40**:

  ```
  <mobrovac> the "extract" field is empty for any page i try to fetch the summary
  for
  <mobrovac> we get that from the MW API
  ```

- **15:08:14**: `thcipriani` is pinged to rollback from `1.30.0-wmf.1` → `1.29.0-wmf.21` since API has
  been pinpointed, `thcipriani` rolls back all wikis on `mwdebug1002` to `1.29.0-wmf.21` and pings
  `mobrovac` and `_Joe_` to test.
- **15:24:19**: RESTBase query⧉ is tried and determined that a rollback does not fix it, but rollback + null edit
  does, caching is involved
- **~15:27**: Help is needed from Mediawiki devs, `ebernhardson` troubleshoots `anomie` is in meeting but
  helps, too
- **15:47:58**:

  ```
  <ebernhardson> hmm, so i can verify at this point that
  ExtractFormatter::filterContent() is the part that's stripping everything, but
  not sure yet what exactly
  <ebernhardson> a full page goes in, and what comes out is just a body tag and an
  html comment
  ```

- **15:58:00**: problem found

```
<ebernhardson> so, after manually removing the <div class="mw-parser-output">
from the beginning, we now get the correct response from
https://en.wikipedia.org/w/api.php?
action=query&format=json&prop=extracts&formatversion=2&redirects=true&titles=Kuipe
belt (because it's cached in memcached from my mwrepl session)
```

```
16:13:08:    <ebernhardson> so random friday options: #1) div.mw-paser-output ->
some other tag? not sure what though #2) revert div.mw-parser-output  #3) Drop
'div' from list of things to strip from text extracts, live with some extra
content that wasn't wanted
16:15:05:    <thcipriani> option 2 is the one that seems like not new code :)
16:32:19:    <mobrovac> anomie, ebernhardson, thcipriani: at this point, I
honestly prefer #2, it seems the safest, the rest can be dealt with during
quieter and more appropriate times
16:33:03:    * anomie submitted https://gerrit.wikimedia.org/r/#/c/353565/ and
https://gerrit.wikimedia.org/r/#/c/353566/ FWIW
16:41:09:    <thcipriani> anomie: so the options to fix would be revert
https://gerrit.wikimedia.org/r/#/c/350634/ or merge these 2
https://gerrit.wikimedia.org/r/#/q/topic:bug/T165161 ?
16:42:37:    <anomie> thcipriani: Yes. Either of those options should also fix
T165115. In both cases additional caching being done by
TextExtracts/MobileFrontend might mask the fix.
16:52:23:    <thcipriani> mobrovac: anomie ok, I'm going to revert
https://gerrit.wikimedia.org/r/#/c/350634/ and run a sync since (a) it's friday
before a week where releng won't be looking deployment stuffs and I don't want
to deploy new code and (b) it looks like it was supposed to be reverted for
wmf.1 anyway https://gerrit.wikimedia.org/r/#/c/352577/ Is there any concern
about a revert breaking anything additional
17:01:58:    <anomie> thcipriani: I don't think reverting in 1.30.0-wmf.1 will
break anything else.
```

- **18:10**: thcipriani@tin: Finished scap: Revert "Wrap parser output in" 4/4 (duration: 19m 13s)
- **18:20**: `legoktm` reports revert not working submits https://gerrit.wikimedia.org/r/#/c/353593/ 🔗 to version the memcached key so that the cache will be cleared
- **18:27**: code live on `mwdebug1002`, but determined not to be working, `MaxSem` explains: `18:35:37: <MaxSem> however, due to reverts of the whole wrapping diff thing, we have corrupt cache`
- **18:43**: `legoktm` creates https://gerrit.wikimedia.org/r/#/c/353596/ 🔗 to clear parser cache and it's deployed, scap catches problem and https://gerrit.wikimedia.org/r/#/c/353597/ 🔗 is created and sync to correct problem
- **19:17**: `thcipriani` syncs https://gerrit.wikimedia.org/r/#/c/353593/ 🔗 to change memcached key

Other service impacts because of the core change

- T165139 - inline <math> editing broken in VE
- T165070 - spurious linter errors reported for <maplink> tags

Underlying cause for both of these is the extra <div> tag around Parsoid's extension output because Parsoid calls the mediawiki API (action=parse) to process extension tags. This has since been addressed in https://gerrit.wikimedia.org/r/#/c/353707/ 🔗 and https://gerrit.wikimedia.org/r/#/c/354050/ 🔗 so that when the core change is redeployed, Parsoid and downstream clients aren't impacted.

Conclusions

- TextExtracts enables a lot of services on which we depend
- TextExtracts can be broken by adding a div to parser output
- There are, evidently, few alarms that catch empty textextract summaries in the API
- This outcome may have been expected judging from a revert that was proposed before the train https://gerrit.wikimedia.org/r/#/c/352577/ 🔗
- The core patch was proposed to be reverted in master as part of https://gerrit.wikimedia.org/r/#/c/352577/ 🔗 but an alternative proposal was to not revert it in master, but to revert it in 1.30.0-wmf.1 instead. But, it appears that the revert didn't happen in wmf.1 before deploy.
- Parsoid uses the mediawiki API to process extension tags and there a number of downstream services that depend on Parsoid: VE, Flow, Content Translation, Android App (via Mobile Content Service), and Linter. This bug would have showed up in the beta cluster as reported in T165139. So, there is clearly a testing gap here.

Actionables

- Better alarm mechanism for empty textextract summaries so they don't languish for 18 hours
- (/me is likely missing lots of subtlety, but armchair-quarterbacks anyway) Adding a div shouldn't break textextract, is there some kind of class/id (e.g., `.mw-ext-textextract-start`) that can be added to parser output that can then be used by textextract that will have the single express purpose of being used by textextract? Adding a div breaking stuff makes the mechanisms by which it is extracting data seem fragile.
- If it is likely that deploying a change will cause breakage (i.e., https://gerrit.wikimedia.org/r/#/c/352577/ 🔗) let deployers/branchers know so that they can backport fixes or reverts before problems manifest.
- Identify how to update the QA tests that are run on the beta cluster so that any potential breakage of VE, etc. is caught before changes are rolled out to production.

Category: Incident documentation

Privacy policy   About   Disclaimers   Code of Conduct   Developers   Statistics   Cookie statement   Mobile view
                   Wikitech

WIKIMEDIA
a project

Powered By
MediaWiki