Agenda

○ Our objectives
○ Progress Q3
○ Tasks Q4
○ Questions and discussion
Our objectives

Deal with wikitext, so you don’t have to™

1. Faithful bidirectional conversion without dirty wikitext diffs
2. Enable new applications and architectural improvements with the HTML+RDFa interface and conversion service
3. Research better templating, widget and diffing solutions
Progress Q3: Ongoing tasks

- Continuous tweaking
  - hacking away at the long tail
    - template encapsulation, foster parenting, nested links, ...
  - informed by testing, bug reports from VE usage

- Continuous deploys
  - mediawiki.org/wiki/Parsoid/Deployments
Progress Q3: Testing infrastructure

- /_rtselser/ endpoint for quick checks
- Parser Tests
  - Reduced false failures, improved selser testing (several automated edit modes + manual edits)
- Round-Trip testing
  - Full-stack testing
  - Moved to hardware (from VMs) for perf. tracking
  - Cassandra storage backend in development by FB OA students
  - OPW intern Be Birchall improved our RT test server code: handlebars templates, new diff library (which doesn’t crash on large diffs), ...
Progress Q3: Images

○ Images and parameters
  ○ basics mostly done
  ○ no shortage of edge cases, big mess in MW
  ○ initiatives to clean up image handling in the longer term
    ■ semantic image formats
    ■ square bounding boxes
    ■ less confusing options, better uniform styling in skin / view
Progress Q3: DOM param editing

○ Parse transclusion parameters to DOM
  ○ originally planned for Q4
  ○ enables visual parameter editing
  ○ parse to DOM ongoing, close to merge
  ○ next step: editing support
Progress Q3: Public API

- [Link](http://parsoid-lb.eqiad.wikimedia.org/)
- Internal users:
  - VisualEditor
  - Flow
  - PDF renderer
  - ContentTranslation
  - Mobile
- Quickly found external users
API users: Kiwix

Offline reader for desktop + mobile

Uses Parsoid for HTML dumps of all wikis, also some non-Wikipedia projects like Wikivoyage
It is requested that an edit be made to this fully protected user page.

This template must be followed by a complete and specific description of the request, so that an editor unfamiliar with the subject matter could complete the requested edit immediately.

Edit requests to fully protected pages should only be used for edits that are either uncontroversial or supported by consensus. If the proposed edit might be controversial, discuss it on the protected page's talk page before using this template. To request that a page be protected or unprotected, make a protection request. When the request has been completed or denied, please add the |answered=yes parameter to deactivate the template.

Respond  Remove request  Done  Needs RS  Unclear/X to Y  Can edit  Misplaced

This is a test

This edit request has been answered. Set the |answered= or |ans= parameter to no to reactivate your request.

Level: Fully protected  Page to be edited:  Answered: 

Response:

(No template response)

Submit
API users: Translation gadget

- Translation gadget Czech <-> Slovak
  - presents machine-translated wikitext to user for copyediting
  - will be generalized by Content translation project

- Google Knowledge Graph
Progress Q3: Services

○ Rashomon revision storage
  ○ Part of larger move to SOA (service-oriented architecture)
  ○ RFC found lots of support at architecture summit
    ■ make new developers more effective
    ■ avoid duplication of effort
    ■ enable independent innovation
    ■ security, monitoring, scaling, testing
    ■ Will be pushed further along by service team
  ○ Not yet in production, but close

○ Debian packaging
  ○ make it easy to install collection of services
Progress Q3: HTML templating

- Originally Q4 research project
- Developed DOM-based library for UI and i18n messages (KnockOff / TAssembly)
  - secure and convenient, yet very fast
- Prototype platform for high-performance content templating
- Cooperation with Flow & Mobile
Progress Q3: Structured Logging

- OPW project: Maria Pacana
  - Generic logging framework
    - App (Parsoid) generates logging events
    - Event stream filtered and dispatched to interested subscribers
  - Structured log format
    - production errors
    - performance metrics
    - wikitext fixup information (GSoC ‘14 proposal)
  - Also useful for development:
    - tracing wikitext parsing actions, debugging, ...
  - No more console.warn, console.error!
Deferred tasks

○ Research: Support switching between HTML and Wikitext within one edit
  ○ hard / low prio in VE

○ Language variant support
  ○ Basic editing support deferred to Q4
  ○ full support questionable

○ Enforce proper nesting of transclusions
  ○ blocked on TemplateData and usage statistics
  ○ GSoC ‘14 proposal (Lintoid/LintTrap) will collect statistics
  ○ probably not before Q1 next fiscal
Tasks Q4

○ Get ready for using Parsoid HTML for all page views
  ○ actually using it stretch goal for Q4/Q1
  ○ in collaboration with VE, Flow, Mobile, Platform
  ○ needs metadata storage service / API
    ■ Rashomon in production
    ■ will be taken over by new services team (Q1)
  ○ requires automated testing to compare PHP parser view and Parsoid views

○ Basic language variant support
  ○ think about longer-term strategy for variant wikis at some point
Tasks Q4

- **HTML templating**
  - support other teams in UI projects
  - continue integration with i18n
    - research conversion of i18n messages to HTML templates using Parsoid
  - research use for content

- **Content widgets**
  - research canned widget solutions for specific tasks (alternative to wikitext hacks)
    - data tables
    - infoboxes
Stretch goals Q4

○ Performance: More efficient template updates
  ○ lower priority than HTML storage
  ○ possible stretch goal for Q4
○ Research: Support switching between HTML and Wikitext within one edit
  ○ hard / low prio in VE
Stretch goals Q4

- Non-Wikipedia projects
  - Parsoid enabled on all public WMF projects
  - biggest issue for editing seems to be labeled section transclusion (primarily Wikisource)
  - potential to fix complexity issues in projects like Wiktionary
  - no time spent on this so far
Thank you!

https://www.mediawiki.org/wiki/Parsoid