



Parsoid

Quarterly review Q2 2013

Agenda

- Our objectives
- Progress Q2 2013
- High-level goals for 2013/14
- Tasks Q3 2013
- Questions and discussion

Our objectives

Make it easy and efficient to view, reuse, and edit content.

1. Convert wikitext content faithfully to semantic HTML+RDFa
2. Support HTML editing without dirty wikitext diffs
3. Use HTML in MediaWiki core
4. Support wikitext editing with Parsoid

Progress Q2 2013

Solid parsing and round-tripping of basically all content without dirty diffs

- Largely met with a combination of several innovative techniques

Line 11:

	voice	=	
	alias	=	Pepe Rodrigo Serrano Gonzales
-	species	=	Muppet [[shrimp]]
	gender	=	Male
	occupation	=	

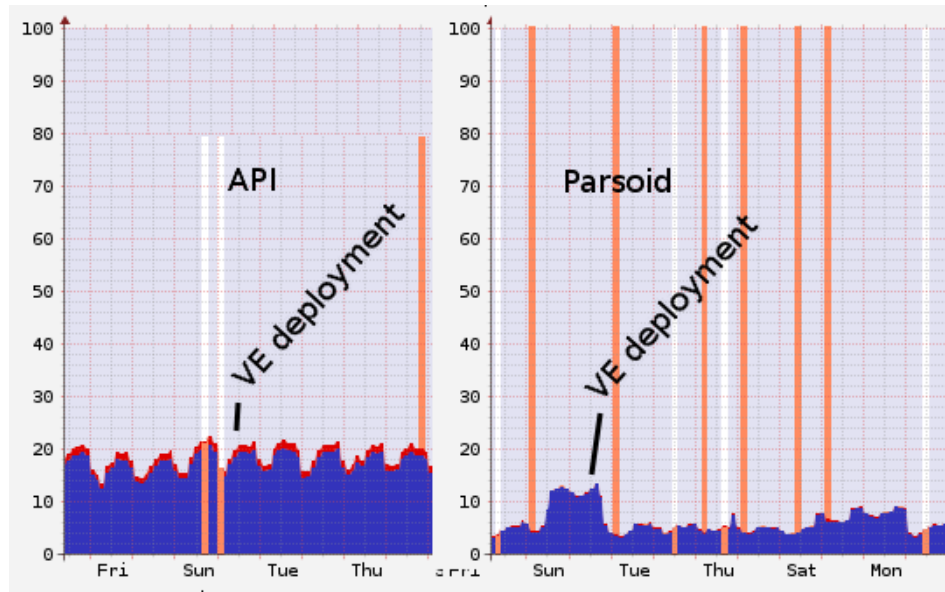
Line 11:

	voice	=	
	alias	=	Pepe Rodrigo Serrano Gonzales
+	species	=	Muppet king prawn
	gender	=	Male
	occupation	=	

Progress Q2 2013

Performance and scaling

- **Met** with pragmatic caching setup, on-edit parsing and expansion reuse
- Paid off to start tracking full load in early June



Progress Q2 2013

Editing support for..

- Templates: Met / Exceeded
- Cite extension: Met
- Tag extensions: Met
- Categories, default sort: Met
- Images and parameters: Partly met, ongoing

Progress Q2 2013

Public HTTP API for VE, Google, bots, offline readers, etc.

- Deferred to Q3 2013

Incremental re-parsing after wikitext edits

- Turned out to not be necessary for release, but idea possibly useful for HTML <-> WP switching in Q3 2013

High-level goals for 2013/14

- Continue to support VE with editing features, bug fixes, etc.
- Start to leverage HTML in MediaWiki core
 - HTML and page property storage (also for Flow?)
 - HTML diffing, basic authorship maps
 - Parsoid HTML for page views (stretch goal)
- Investigate HTML-based templating

See <https://www.mediawiki.org/wiki/Parsoid/Roadmap>

Tasks Q3 2013

- Image editing refinements [straightforward]
- Provide public HTML API [straightforward]
- Research: Language variant support [hard, Q3-Q4?]
- Research: Support switching between HTML and Wikitext within one edit [hard, Q3-Q4 2013?]

Tasks Q3 2013

- HTML / Wikitext compound storage; support Flow [medium, Q3-Q4 2013]
- Enforce proper nesting of transclusions [hard, Q3-Q4 2013]
- Testing infrastructure improvements [straightforward, Q3-Q4 2013]
- Performance: More efficient template updates [straightforward]

Other tasks on the horizon

- Parse most transclusion parameters to DOM once type info is available [medium, likely Q4 2013]
- HTML-only wiki support [hard, Q4 2013 - Q2 2014]
- Non-Wikipedia projects [likely hard, Q4 2013 - Q1 2014?]
- Research DOM-based templating [hard, Q4 2013 - Q1 2014]
- Use Parsoid HTML for all page views [hard, stretch goal, Q2 2014?]

Questions?