Rich text & parsing

EDITING THE FUTURE

THIS! IS! PARSER!

- MediaWiki's current parser is split between two components:

- Well-defined Preprocessor (template expansion, parser functions & tag hooks, headings, comments)

- Poorly-defined block & inline processing in Parser class

THIS! IS! PARSER!

- Preprocessor is fast and reversible

- Later parsing throws away data, makes irreversible transformations

- -> we cannot easily reverse HTML to wikitext, or identify source wikitext for all HTML elements

Does better editing **need** a new parser?

Hopes

- **Immediate, fine-grained editing**

- jump straight from page view into a table cell, template parameter, reference contents, etc.

- Need enough annotations from the parser on **real output** that we can identify the part of the source that needs changing

Existing editor work

- FCK - uses core parser

- Wikia RTE - modifies core parser

- WikiHow - site-specific editor

- Twoonix - custom parser

- Wysiftw - custom parser -> section editing

- InlineEditor - custom -> para editing

Hopes II

- **Content portability is a key Wikimedia goal**

- Tighter specification of later parser steps will make our content more portable

- Will also make templates more portable

- Taking emphasis off source-level editing encourages future features to use existing function syntax

Read my lips:
No New Syntax

- Refine, don't reimagine

- Decoupled: test on live sites via gadgets

- Inclusive: make easy to try, easy to report issues

- Widespread: aggressively do batch testing to confirm things work on complex pages

- Scariest problem: open...close templates

- Try to find a way to leave these as-is and make the parser & editor smart enough to handle it

- If can't, a clean migration path is **key to deployment**

What's happening now?

- Brion -- more experiments with JavaScript parsing w/ PEG definitions & custom AST->HTML transform

- Trevor -- experiments with visual editor surface

- Neil -- experiments with another JS PEG-based parser

- You -- ???

Timeline

- May 2011 -- kickoff here!

- ... experimentation in gadgets...

- August 2011 -- Wikimania demos

- ... opt-in early tests in gadgets...

- November 2011 -- Larger betas & opt-in testing

- ... deployment & migration TBD

Things to do right now?

- Fiddle with ParserPlayground gadget on mediawiki.org

- Keep experimenting with Wysiftw, InlineEditor, Wikia RTE, and any custom parsers - mwlib, sweble, etc

- Be aggressive about finding real-world test cases: interesting templates on real pages that fail on existing parsers

http://mediawiki.org/wiki/Future