Wikisource technical infrastructure
What we have done and what we could do
II

Thomas Pellissier Tanon
User:Tpt
@Tpt93

Wikisource Conference 2015

Wikisource ?

- Current state:
    - 4 millions of pages
    - 2.1 millions of proofread pages
    - 600 active editors (> 5 edits)

Wikisource ?

- Current state:
    - 4 millions of pages
    - 2.1 millions of proofread pages
    - 600 active editors (> 5 edits)
- Strong issues:
    - books not easily accessible
    - no real bibliographic database
    - contributing is quite difficult

Wikisource technical infrastructure

- MediaWiki

Wikisource technical infrastructure

- MediaWiki
- but with custom extensions like ProofreadPage

Wikisource technical infrastructure

- MediaWiki
- but with custom extensions like ProofreadPage
- developed and maintained by volunteer contributors and a few GSoC projects

Outline

What we have done

Wsexport

- Is a "magic" export tool

Wsexport
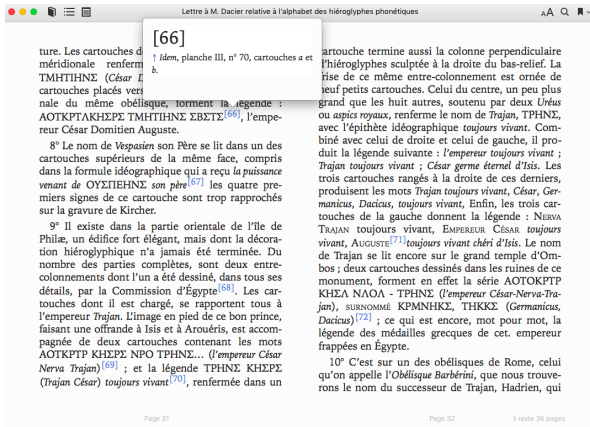
- Is a "magic" export tool
- adapted to Wikisource needs

Wsexport

- Is a "magic" export tool
- adapted to Wikisource needs
- ePub is the base format
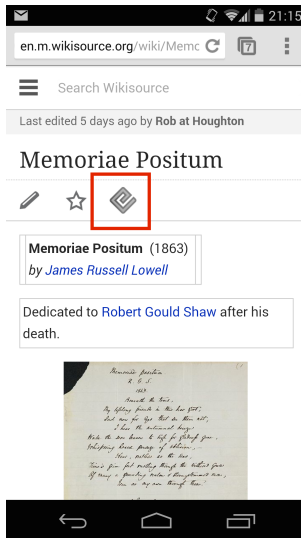
- Migrated on Wikimedia Tools labs at `https://tools.wmflabs.org/wsexport`

Wsexport

- Migrated on Wikimedia Tools labs at `https://tools.wmflabs.org/wsexport`
- Integrated in the UI of most of Wikisources

Wsexport

- Migrated on Wikimedia Tools labs at `https://tools.wmflabs.org/wsexport`
- Integrated in the UI of most of Wikisources
- 48,000 exports in October 2015
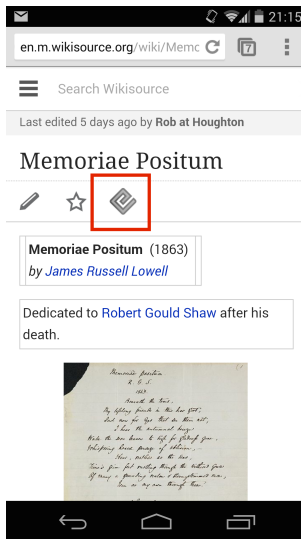
Wsexport

- Migrated on Wikimedia Tools labs at `https://tools.wmflabs.org/wsexport`
- Integrated in the UI of most of Wikisources
- 48,000 exports in October 2015
- Supports PDF, mobi...
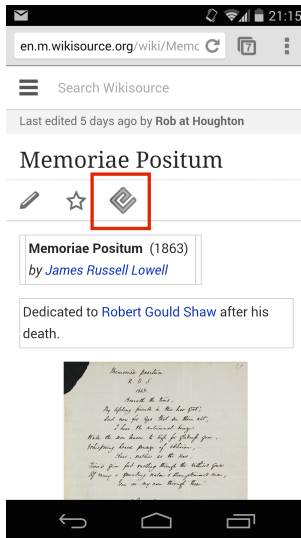
Wsexport

- Migrated on Wikimedia Tools labs at `https://tools.wmflabs.org/wsexport`
- Integrated in the UI of most of Wikisources
- 48,000 exports in October 2015
- Supports PDF, mobi...
- ePub 3 is the default format
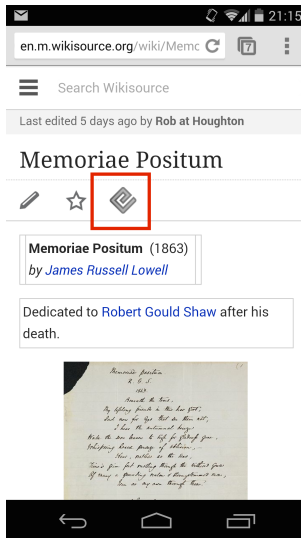
Refactoring of ProofreadPage

Goals:

- More maintainable code
- Use new MediaWiki features (`ContentHandler`...)

Refactoring of ProofreadPage

Goals:

- More maintainable code
- Use new MediaWiki features (`ContentHandler`...)
- Better performances
- Less breakages

Refactoring of ProofreadPage
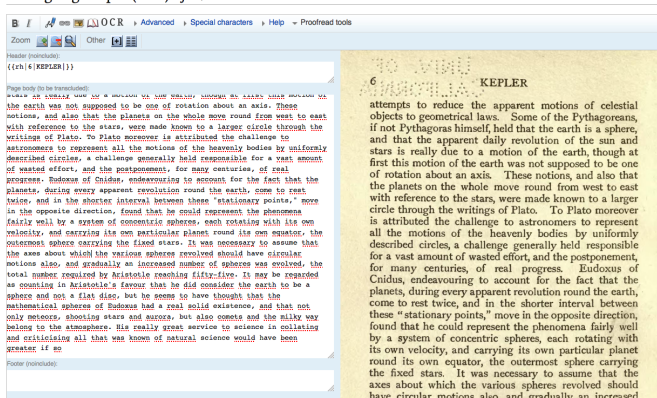
Goals:

- More maintainable code
- Use new MediaWiki features (`ContentHandler`...)
- Better performances
- Less breakages
- Content validation

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP

Editing Page:Kepler(1920).djvu/12

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP
- Try to have not too badly architectured code

Editing Page:Kepler(1920).djvu/12

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP
- Try to have not too badly architectured code
- Automated tests

Editing Page:Kepler(1920).djvu/12

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP
- Try to have not too badly architectured code
- Automated tests
- JSON encoding of Page: pages in API

Editing Page:Kepler(1920).djvu/12

Interproject sidebar

IA-Upload

An Internet Archive to Commons import tool
(http://tools.wmflabs.org/ia-upload)

BUB

An import tool for Internet Archive from Google Books and other sources
(https://tools.wmflabs.org/bub/)

What we could do

Some ideas...

A Wikisource contributors survey done in Fall 2013 with 251 answers

What do you think are the core priorities for the Wikisource community?
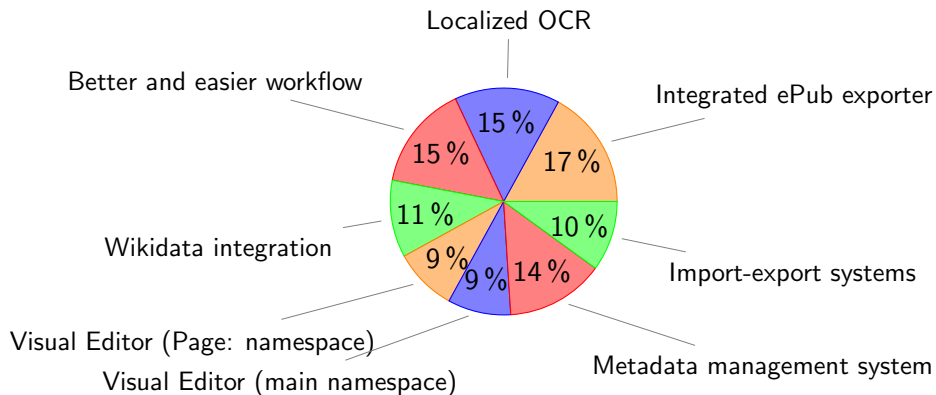
VisualEditor support

- Improve Parsoid rendering of Wikisource content

VisualEditor support

- Improve Parsoid rendering of Wikisource content
- Support tags used on Wikisource (pages, poem, section...)

VisualEditor support

- Improve Parsoid rendering of Wikisource content
- Support tags used on Wikisource (`pages`, `poem`, `section`...)
- Custom interface for Page: pages

- Custom edit interface for Page: and Index: pages

Editing Page:La Brière - Ch...   Next

```
<noinclude><pagequality level="4"
user="Maniak" /><div class="pagetext">


</noinclude>{{petit titre|I}}


{{petit titre|CHAMPOLLION ENFANT}}


Le génial savant qui, par une divinatrice
intuition, découvrit, au commencement de ce
siècle, la clé perdue de l'écriture égyptienne,
déchirant soudain les voiles d'une histoire
mystérieuse, exhumant des entrailles du passé
```

Mobile support

- Custom edit interface for Page: and Index: pages
- We should have a nice UI for both browsing and editing

Mobile support

- Custom edit interface for Page: and Index: pages
- We should have a nice UI for both browsing and editing
- Future: gamification?

Exportation of content

- Improve Wsexport performances

Exportation of content

- Improve Wsexport performances
- Use Parsoid

Exportation of content

- Improve Wsexport performances
- Use Parsoid
- Nice book browsing interface + OPDS?

Exportation of content

- Improve Wsexport performances
- Use Parsoid
- Nice book browsing interface + OPDS?
- Integrated inside of Wikimedia infrastructure?

Some other ideas

- Use Wikidata as much as possible

Some other ideas

- Use Wikidata as much as possible
- Gamification (capcha...)

Some other ideas

- Use Wikidata as much as possible
- Gamification (capcha...)
- . . .

Conclusion

What we need now:

- Stronger interwiki collaboration

Conclusion

What we need now:

- Stronger interwiki collaboration
- People to build the Wikisource of tomorrow

Conclusion

What we need now:

- Stronger interwiki collaboration
- People to build the Wikisource of tomorrow
- Stronger support from the Wikimedia movement

Thanks a lot for your attention!