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

Thomas Pellissier Tanon
User:Tpt
@Tpt93

Wikisource Conference 2015
Current state:
- 4 millions of pages
- 2.1 millions of proofread pages
- 600 active editors (> 5 edits)
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
1 What we have done

2 What we could do
What we have done
- **Wsexport**

- **Is a "magic" export tool**

---

*Letter to M. Didier relative to the alphabet of the hieroglyphs phoneticals*

>... Les cartouches de la méridionale renferment: ΤΜΗΤΙΗΝΣ (César I) cartouches placés verticalement du même obélisque, forment la légende: ΑΟΤΚΡΤΛΔΕΙΨΕΙΣ ΤΜΗΤΙΗΝΣ ΣΒΕΣΤΣ, l'empereur César Domitien Auguste.

8° Le nom de Vespasien son père se lit dans un des cartouches supérieurs de la même face, compris dans la formule idéographique qui a reçu la puissance venant de ΟΥΣΙΕΠΗΝΣ son père[67] les quatre premiers signes de ce cartouche sont trop rapprochés sur la gravure de Kircher.

9° Il existe dans la partie orientale de l'île de Philae, un édifice fort élegant, mais dont la décoration hiéroglyphique n'a jamais été terminée. Du nombre des parties complétées, sont deux entrecollonements dont l'un a été dessiné, dans tous ses détails, par la Commission d'Égypte[68]. Les cartouches dont il est chargé, se rapportent tous à l'empereur Trajan. L'image en pied de ce bon prince, faisant une offrande à Isis et à Arouérés, est accompagnée de deux cartouches contenant les mots: ΑΟΤΚΡΠΤ ΚΗΠΕΡ ΝΡΟ ΠΡΗΝΣ (l'empereur César Nerva Trajan)[69] et la légende ΤΡΗΝΣ ΚΗΠΕΡ (Trajan César) toujours vivant[70], renfermés dans un

---

Thomas Pellissier Tanon  Wikisource technical infrastructure 2  Wikisource Conference 2015  6 / 20
Wsexport

- Is a "magic" export tool
- adapted to Wikisource needs
Is a "magic" export tool
adapted to Wikisource needs
ePub is the base format
Wsexport

- Migrated on Wikimedia Tools labs at https://tools.wmflabs.org/wsexport
Migrated on Wikimedia Tools labs at https://tools.wmflabs.org/wsexport

Integrated in the UI of most of Wikisources
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

Memoriae Positum

(1863)

by James Russell Lowell

Dedicated to Robert Gould Shaw after his death.
- 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...
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

Thomas Pellissier Tanon

Wikisource technical infrastructure 2

Wikisource Conference 2015 9 / 20
Refactoring of ProofreadPage

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

attempts to reduce the apparent motions of celestial objects to geometrical laws. Some of the Pythagoreans, if not Pythagoras himself, held that the earth is a sphere, and that the apparent daily revolution of the sun and stars is really due to a motion of the earth, though at first this motion of the earth was not supposed to be one of rotation about an axis. These notions, and also that the planets on the whole move round from west to east with reference to the stars, were made known to a larger circle through the writings of Plato. To Plato moreover is attributed the challenge to astronomers to represent all the motions of the heavenly bodies by uniformly described circles, a challenge generally held responsible for a vast amount of wasted effort, and the postponement, for many centuries, of real progress. Eudoxus of Cnidus, endeavouring to account for the fact that the planets, during every apparent revolution round the earth, come to rest twice, and in the shorter interval between these “stationary points,” move in the opposite direction, found that he could represent the phenomena fairly well by a system of concentric spheres, each rotating with its own velocity, and carrying the own particular planet round its own equator, the outermost sphere carrying the fixed stars. It was necessary to assume that the axes upon which the various spheres revolved should have circular motions also, and gradually the number of spheres was evolved, the total number required by Aristotle reaching fifty-five. It may be regarded as pointing in Aristotle’s favour that he did consider the earth to be a sphere and not a flat disk, but he seems to have thought that the mathematical spheres of Eudoxus had a real solid existence, and that not only meteors, shooting stars and aurora, but also comets and the milky way belong to the atmosphere. His really great service to science in collating and criticising all that was known of natural science would have amounted if we...
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
The ruins of Susa now being excavated by archaeologists have already yielded important results. He was the first to find inscriptions in the ruins of Babylonia, indicating the presence of important monuments of the Babylonian and Elamite kings, to their capital, Susa. When recovering from the ruins of Susa these remains were first found by the stele of Narâm-Sîn on the Tigris-Euphrates valley nearly 6,000 years ago. This code is the oldest collection of public laws existing in Babylonia 4,000 years ago.

(Read on... November's featured text)

or see all featured texts.

The current Community collaboration

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

**Figure:** What do you think are the core priorities for the Wikisource community?

- Integrated ePub exporter: 17%
- Localized OCR: 15%
- Better and easier workflow: 15%
- Wikidata integration: 11%
- Visual Editor (Page: namespace): 9%
- Visual Editor (main namespace): 9%
- Import-export systems: 14%
- Metadata management system: 10%
VisualEditor support

- Improve Parsoid rendering of Wikisource content
VisualEditor support

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

Thomas Pellissier Tanon

Wikisource technical infrastructure 2

Wikisource Conference 2015
VisualEditor support

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

---

I

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 toute une science nouvelle, laquelle dépassa presque aussitôt les balbutiants essais, et atteint déjà de notre temps sa maturité pleine, Jean-François Champollion, l’une des gloires les plus pures et les plus universellement reconnues de notre panthéon national, n’est inconnu d’aucun...
Mobile support

- Custom edit interface for Page: and Index: pages

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!