



Language Engineering Team

Wikimedia Foundation

Quarterly Review September 3 2014



Agenda

Introductions and Team Overview

Q4 2013-14 Project Goals and Accomplishments

Q1 2014-15 Project Goals, Accomplishments, Timelines

Inter-team dependencies

Reusable i18n and L10n tools

Looking to the Future - What's on our roadmap?

Questions & Answers

Introductions



Alolita Sharma

Director of language
engineering



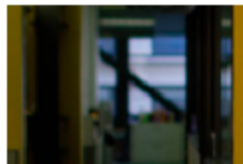
Amir E. Aharoni

Software engineer



David Chan

Software engineer



Joel Sahleen

Software engineer



Kartik Mistry

Software engineer



Niklas Laxström

Software engineer



Pau Giner

User experience
designer



Runa Bhattacharjee

QA and outreach
coordinator



Santhosh Thottingal

Senior software
engineer

Our Team Strengths

- Language technology
- Standards
- Engineering process maturity; Agile
- Innovation
- Cross-team collaboration
- Internal: Wikimedia
- External: Red Hat, Google, Mozilla, jQuery, other open source projects
- Transparent
- Diverse
- Multilingual
- Global
- Friendly

Team Overview

Our Mandate

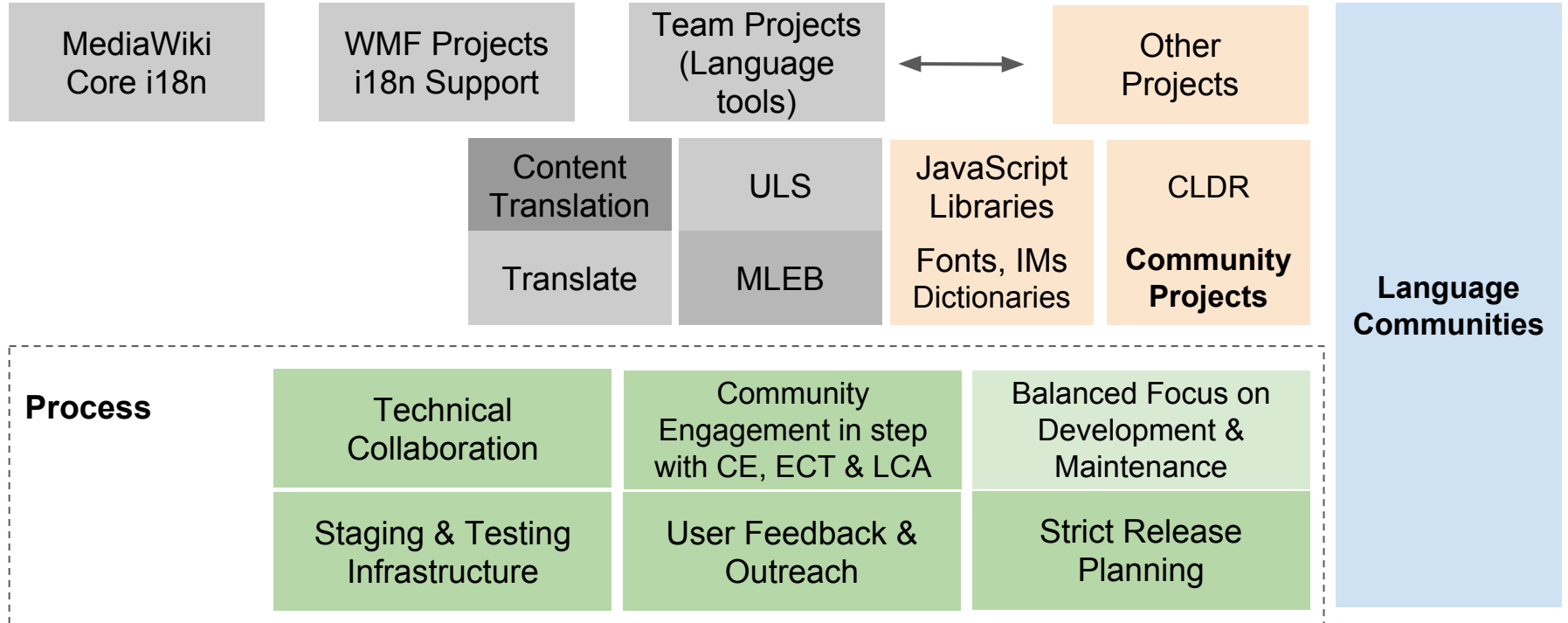
1. Provide language support for all Wikimedia websites
2. Provide internationalization and localization expertise to other Wikimedia engineering teams
3. Develop reusable, standards based internationalization and localization tools and be exemplary open source citizens

Our Team's Impact is Global

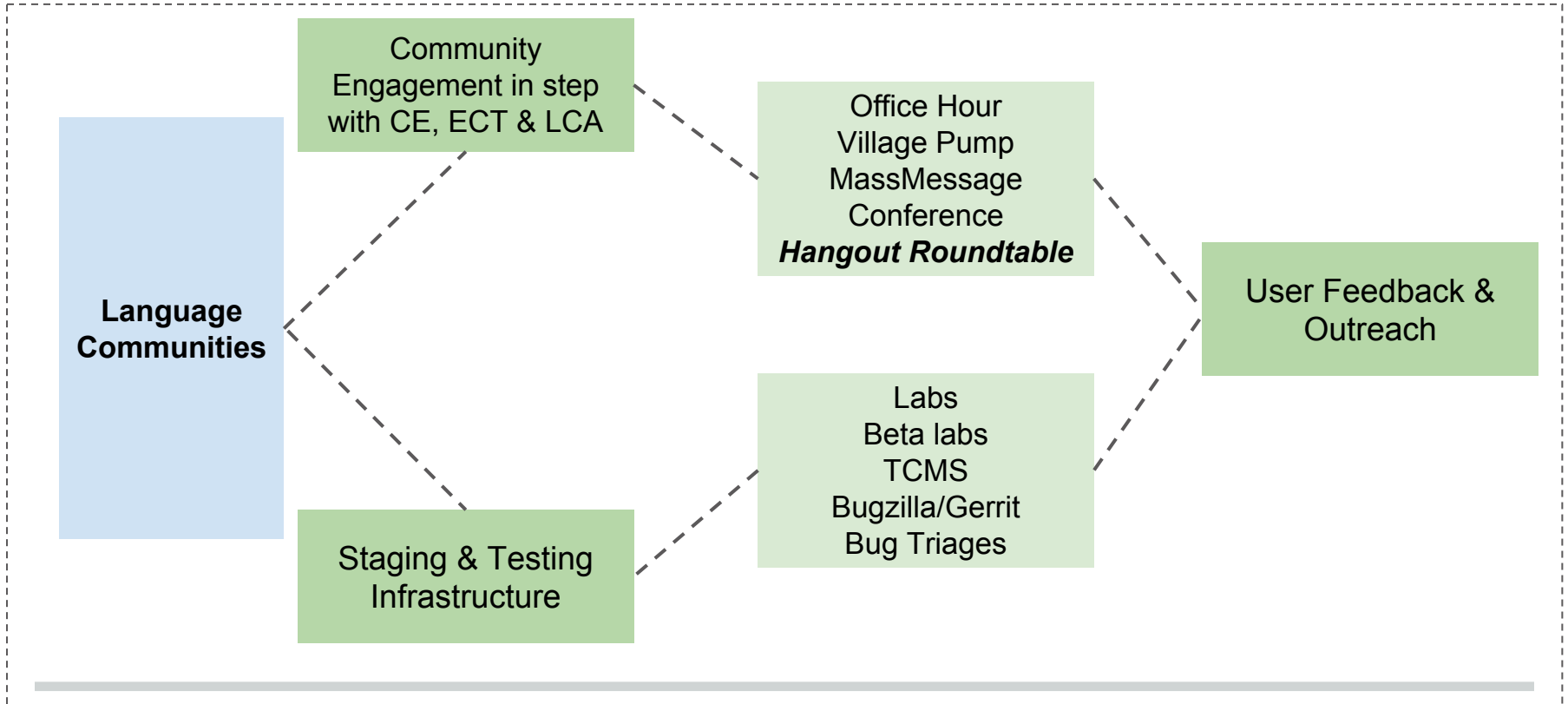
1. Our multilingual users across the world
 2. MediaWiki developer community
 3. Wikimedia Foundation Product
-

April-June 2014: Engineering Summary

Projects

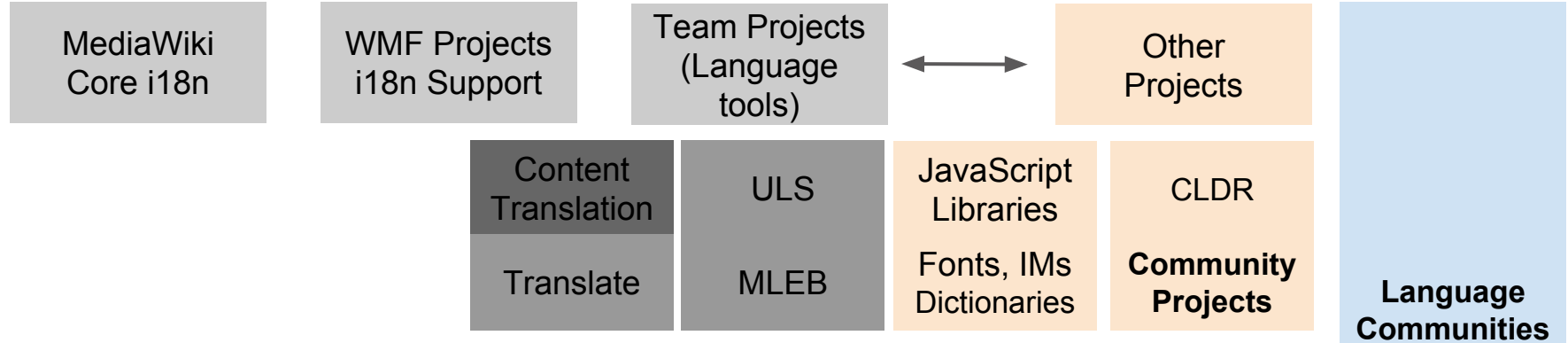


Engineering Summary

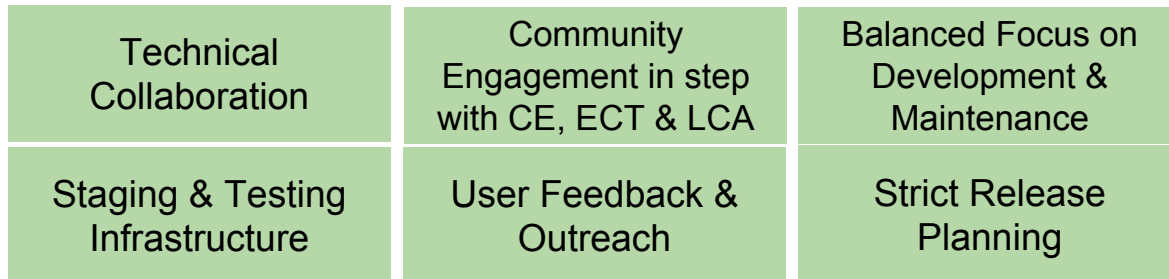


July-Sep 2014: Engineering Summary

Projects



Process





Publish translation

Mar Báltico

español [view page](#)

El **mar Báltico** (del *latín* *Mare Balticum*) es un **mar** interior de **agua salobre** del norte de **Europa** abierto al **mar del Norte** y, finalmente, al **océano Atlántico** a través de los estrechos de **Kattegat** y **Skagerrak**. Los países que lo rodean son (empezando por la **península Escandinava** y siguiendo en sentido horario): **Suecia**, **Finlandia**, **Rusia** (**óblast de Leningrado** y **Kaliningrado**), **Estonia**, **Letonia**, **Lituania**, **Polonia**, **Alemania** y **Dinamarca**.

Su **superficie** es de 432.800 km² e incluye dos grandes golfos: el **golfo de Finlandia**, entre el sur de este país y **Estonia**, y el **golfo de Botnia**, entre la costa oriental de Suecia y el occidental de Finlandia. Geológicamente es muy joven: sólo existe desde el VI milenio a.C. Es muy poco profundo (la media es de 57 m; la **profundidad** máxima es de 459 m al norte de la isla sueca de **Gotland**), lo que, unida a la poca apertura al **océano**, hace la renovación de las aguas muy lenta y favorece los problemas de **contaminación**. Las **mareas** son de muy pequeña amplitud.

El mar Báltico es también, el mayor depósito de **ámbar** del mundo y, además, su calidad es de las mejores: de los veinte

Mar Báltico

català

El **mar Báltic** (del *llatí* *Mare Balticum*) és un mar interior d'aigua salobre del nord d'Europa obert al mar del Nord i, finalment, a l'oceà **Atlàntic** a través dels estrets de Kattigat i Skagerrak. Els països que ho envolten són (començant per la península Escandinava i seguint en sentit horari): **Suècia**, Finlàndia, Rússia (óblast de Leningrad i Kaliningrado), **Estònia**, **Letònia**, **Lituània**, **Polònia**, **Alemanya** i **Dinamarca**.

Search for a word



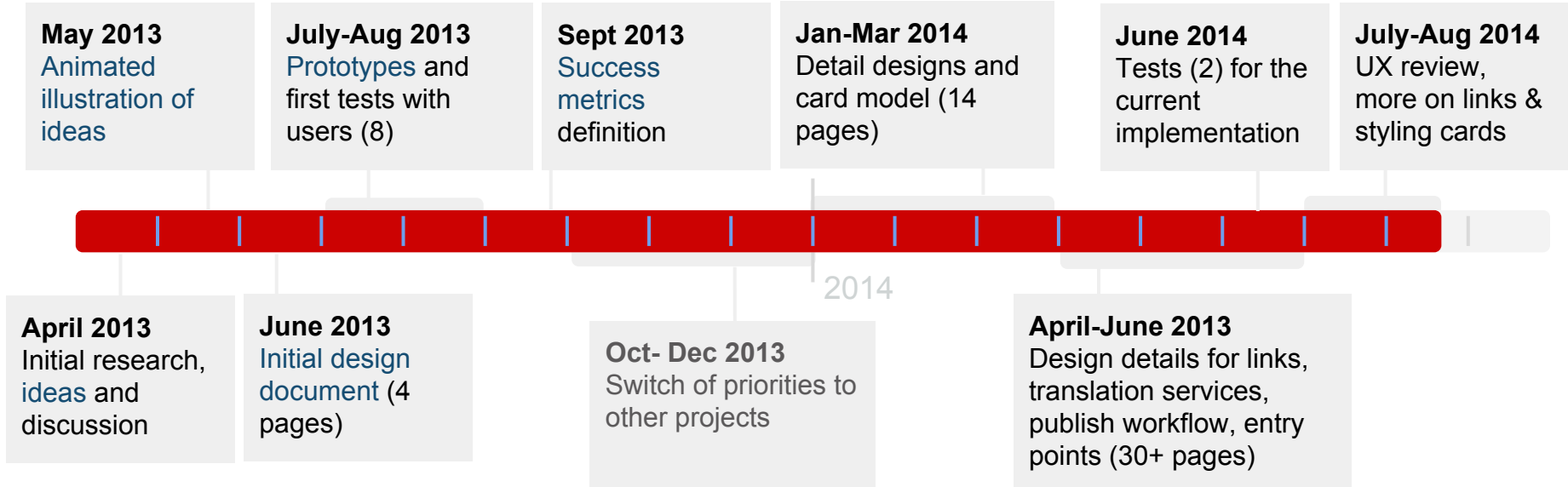
Link	español
Link	català
Mar	
+ Add link	

Content Translation (CX)

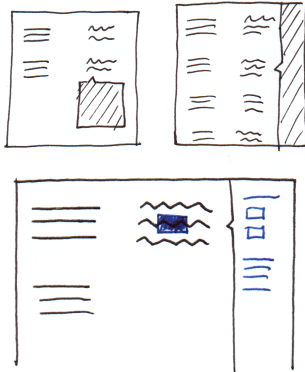
April-June 2014: CX Features

- Tool system basic frontend and implementation
 - Dictionaries
 - Link adaptation
 - Machine translation
 - References
 - Beta environment preparation
 - Server infrastructure and performance
-

CX Design Evolution Timeline



April-June 2014: CX Design Evolution



WIKIPEDIA

Translation center English to Español 0% 3 Review and publish

Ludwig Mies van der Rohe

Ludwig Mies van der Rohe (born **Marie Ludwig Michael Mies**; March 27, 1898 – August 19, 1969) was a German-American architect.^[1] He is commonly referred to, and was addressed, as Mies, his surname. He served as the last director of the Bauhaus. Along with Le Corbusier, Alvar Aalto, and Frank Lloyd Wright, he is widely regarded as one of the pioneering masters of modern architecture.

Early career

Mies worked in his father's stone-carving shop and at several local design firms before he moved to Berlin, where he joined the office of interior designer Bruno Paul.^[2] He began his architectural career as an apprentice at the studio of Fritz Höber from 1909 to 1912, where he was exposed to the current design theories and to progressive German culture, working alongside Walter Gropius and to Gropius, who were later also involved in the development of the Bauhaus. Mies served as construction manager of the Embassy of the German Empire in Saint Petersburg under Behrens.^[3]

Suggestions

- Ludwig Mies van der Rohe (nacido **Marie Ludwig Michael Mies**, 27 marzo 1898 hasta 19 agosto 1969) fue un arquitecto germano-estadounidense [1] El se conoce comúnmente, y se dirigió, como Mies, su apellido. From Google translator
- Ludwig Mies van der Rohe (nacido **Marie Ludwig Michael Mies**, 27 marzo 1898 hasta 19 agosto 1969) fue un arquitecto germano-estadounidense [1] El se conoce comúnmente, y aboró, como Mies, su apellido. From Microsoft translator

Use source text Clear suggestion

WIKIPEDIA

Create account Log in

All translations Saved 2 minutes ago Publish translation

Traffic light

English View article

신호등

Korean

신호등 (信號燈)은 철도나 도로에 있어서 교통 안전 확보 또는 교통의 흐름을 원활히 하기 위해서 진행 방향 등의 신호를 나타내는 장치이다.

Traffic lights, also known as traffic signals, traffic lamps, signal lights, stop lights and robots,^[1] and also known technically as traffic control signal^[2] are signalling devices positioned at road intersections, pedestrian crossings and other locations to control competing flows of traffic. Traffic lights were first installed in 1868 in London^[3] and are now used all over the world.

Traffic lights alternate the right of way accorded to road users by displaying lights of a standard color (red, orange/amber, and green) following a universal color code. In the typical sequence of color phases:

the green light allows traffic to proceed in the direction denoted, if it is safe to do so
the yellow/amber light denoting prepare to stop short of the intersection, if it is safe to do so
the red signal prohibits any traffic from proceeding

Definition

Signal (noun)

A sign made to give notice of some occurrence, command, or danger, or to indicate the start of a concerted action.

* 신호를 시간

Signal (verb)

To indicate
신호를 보낸다

April-June 2014: CX UX Goals

- Make it easy to start and complete a translation
 - Main entry point, start translating, communicate waiting, and publishing workflow.
- Greater control of links, and translations
 - Support for multi-language cards and translation services.

New translation

From: English

To: Korean

[Create translation](#)


+ Add translation

er page [Discussion](#) [Read](#) [Edit](#) [View history](#) [Actions](#)

Congratulations, the page has been published

You can click "edit" to keep improving the page. Make sure it reads natural.

[Ok](#)

	Link Signal English
	Link 신호를 Korean
	+ Add link

Automatic translation

Not available for Galician

Use source text

Google translate (Portuguese)

Don't use automatic translation

April-June 2014: CX UX Goals

- Confirm our assumptions with users
 - Creating a new article is a natural usecase
 - Fluency on a 3-column paragraph-by-paragraph editor
 - Good automatic translation speeds up the process
 - Advanced scenarios (collaborative editing, move paragraphs, edit complex elements...) are not project goals
-

July-Sep 2014: CX UX Goals

- Get a more robust and fluent experience
 - Review UX and polish details
 - Improve link and styling support
- Measure our success:
 - Feedback from users (testing and reported)
 - Check resulting articles (and post-creation activity)
 - Analytics (currently: total articles)

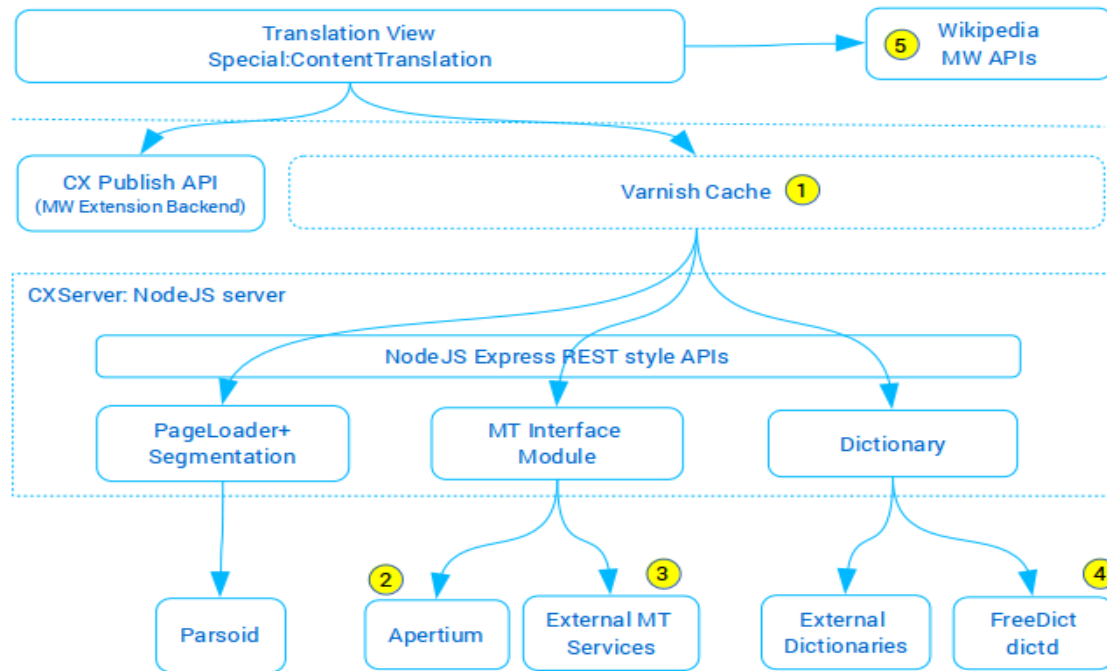
traffic signals,
top light and
technically as traffic
ing devices posi-
pedestrian cross-

신호등(信號燈)은 철도나 도로에 있
교통 안전 확보 또는 교통의 흐름을
하기 위해서 진행·정지등의 신호를



115. [Michael Freedman](#)
116. [Jean-Christophe Yoccoz](#)
117. [Elon Lindenstrauss](#)
118. [Andrei Okounkov](#)
119. [William Thurston](#)
120. [Liubov Popova](#)
121. [Pad thai](#)

CX Technology Architecture



April-June 2014: CX Technology

Architecture and components

- Architecture discussed and improved with Gabriel, Ops
- Node.js is used as server technology, powered by Parsoid, cache by Varnish
- Open source MT Apertium chosen as first MT engine, worked well for Spanish<->Catalan pair

https://www.mediawiki.org/wiki/Content_translation/Technical_Architecture

April-June 2014: CX Technology

Strategy

- Language tools designed as plugins, potentially enabling community to write translation support tools
 - Challenged by availability of free licensed language tools (MT, Dictionaries)
 - UX tries to avoid over-dependence on MT and fills the gap with all other possible tools (links, references)
-

July-Sep 2014: CX Technology

Ongoing Activity

- Working with Ops team to get deployed in production
 - Currently available in beta cluster
 - Improving Apertium backend to be a stable and scalable system
 - Currently using Apertium as a server - APY
-

July-Sep 2014: CX Technology

September 2014 Release Plan

- Optimizing modules and bandwidth usage for API requests
 - Improving template, links support
 - Formatting support - minimal formatting toolbar
 - Code refactoring, cleanup
 - UX fixes and improvements
-

July-Sep 2014: CX Technology

Add supported language pairs

- **In labs, for limited internal testing:**
 - Release all languages supported by Apertium to collect limited feedback
 - **In beta, for testing by users and feedback collection:**
 - Release
 - Catalan to Spanish
 - Portuguese <-> Spanish
 - Portuguese <-> Catalan
-

July-Sep 2014: CX User Feedback

- Deployed to beta on July 17
- Supporting Spanish to Catalan language pair
- Used daily since then for publishing articles
- Continued development: template support, bug fixing, more languages
- Updates can be found at: https://www.mediawiki.org/wiki/Content_translation/Published_pages

138

Published
Articles

17

Unique
Translators

Select language



Language search

Common languages

català

नेपाली

हिन्दी

ଓଡ଼ିଆ

English

भोजपुरी

অসমীয়া

தமிழ்

اردو

मराठी

বাংলা

తెలుగు

मैथिली

ગુજરાતી

ಕನ್ನಡ

മലയാളം

Worldwide

Universal Language Selector



Display settings



Input settings

Universal Language Selector: Status

- ULS Webfonts continue to be disabled by default.
 - Enabled default Webfonts delivery for English, Hebrew Wikisources and Divehi wikis by community request.
 - Analytics of font usage, user preferences for Webfonts are collected. Not analyzed. Prioritization pending.
 - Language selection by anonymous users continues to be unsupported.
-

Gives the user several ways to select a language and to adjust language settings

All Untranslated Outdated Translated ...

Filter list

MediaWiki:Uls-desc/ml ▾



Gives the user several ways to select a language and to adjust language settings

ദാഷ തിരഞ്ഞെടുക്കുന്നതിനും സജ്ജീകരിക്കുന്നതിനുമുള്ള സംവിധാനം

Save translation

Skip to next



Press "ALT-SHIFT-S" to save or "ALT-SHIFT-D" to skip to next message or "ALT" for other shortcuts.

Extension description

[/ Edit documentation](#)

Suggestions

In other languages

सदस्य को भाषा चुनने और भाषा विकल्प बदलने के कई तरीके देता है हिन्दी

Need more help? [Ask for more information](#)

Languages ഭാഷകൾ ✓ Translated [/ Edit](#)

Language settings ഭാഷാ സജ്ജീകരണങ്ങൾ ✓ Translated [/ Edit](#)

Language changed from \$1 ഭാഷ മാറിയിരിക്കുന്നു. പഴയ ഭാഷ \$1 ✓ Translated [/ Edit](#)

More language settings കൂടുതൽ ഭാഷാസജ്ജീകരണങ്ങൾ

Compact language links ഒതുക്കമുള്ള ഭാഷാ കണ്ണികൾ

Show a shorter version of the language list, with just the languages that are more relevant to you. താങ്കളെ സംബന്ധിച്ച് പ്രസക്തമായ ഭാഷകൾ മാത്രം ... ✓ Translated [/ Edit](#)



Translate: Status

Powers translation of interface messages, newsletters, and documentation.

- Our own translation memory
 - Started migrating from Solr to Elasticsearch
 - Ongoing bug fixes
 - Prevent data loss on session expiry
 - GSoC project to migrate old translated pages, mark source language and other improvements to page translation
-

July-Sep 2014: Translate Todos

- Translation memory
 - Performance
 - Suitability for CX
 - Page translation
 - Replication lag issues
 - JobQueue interaction
 - Does not integrate with Parsoid and VE
-

Reusable Tools for Internationalization and Localization

Reusable Tools for Internationalization and Localization

- Software development, maintenance support for translatewiki.net, the platform for UI translations for open source software projects including MediaWiki
 - **MediaWiki Language Extension Bundle (MLEB)** to use our internationalization and localization extensions, tools and resources such as fonts, input tools
 - **Project Milkshake**: standalone jQuery language libraries which can be used in other projects
 - **CLDRPluralRuleParser** - Plural rules processing for languages based on CLDR
-

Reusable Tools for Internationalization and Localization

jquery.ime - Open Source Input Method library

- Input methods are constantly getting added and maintained
- Used in Koha, Firefox OS, Android indic keyboard app, Firefox/Chrome Browser extensions
- Fixing the input method related issues in VE (In progress by David Chan)

Plans

- UI standardization of input widgets in MediaWiki opens up opportunity for better UI of IMEs
- IME integration with CX is planned

86
languages

143
input
methods

Reusable Tools for Internationalization and Localization

jquery.i18n - Open Source UI internationalization library

- Port of MediaWiki localization system
 - Used as the i18n library of ULS
 - Used in standalone VisualEditor library for internationalization
 - JSON based message formatting first defined in jquery. i18n is the MediaWiki localization file format now.
 - Unifying the MediaWiki JavaScript i18n system and this library is a thought for the future
-

Reusable Tools for Internationalization and Localization

jquery.uls - Open Source advanced language selector and language information library

- Used in Universal Language selector, Translate, ContentTranslation, VisualEditor, Wikidata, translatewiki.net
 - Also used in standalone MediaWiki independent VisualEditor
 - Currently in maintenance mode
-

Reusable Tools for Internationalization and Localization

jquery.webfonts - Open Source font embedding library

- Used in Universal Language Selector for serving Webfonts
 - Currently in maintenance mode
-

Reusable Tools for Internationalization and Localization

CLDRPluralRuleParser: Open Source CLDR plural rule system

- A Javascript implementation of CLDR plural rule system for languages
 - Powers plural rule handling of MW JavaScript side and jquery.i18n
 - Recently relicensed and adopted by jQuery Globalize and moment.js (which is used in MediaWiki for human readable time labels) - An ongoing collaboration between MediaWiki and jQuery projects
-

Inter-Team Dependencies

Inter-Team Dependencies

We depend on other teams to support us for language tools development

- Product Needs
 - Customer needs - better feedback, surveys
 - Impact analysis
 - UX
 - User Experience testing and analysis
 - Analytics and Research
 - Metrics for measuring impact
 - Community
 - Community Engagement (CE)
 - Engineering Community Team (ECT)
 - Legal and Community Advocacy (LCA)
-

Inter-Team Dependencies

We depend on other teams to support us for language tools development

- Release Engineering
 - Performance testing, metrics and consultation
 - Beta labs support
 - Code Review, Deployments
 - QA
 - Cross browser testing
 - Multi-language test environments
 - Integration testing
 - Ops
 - Hardware provisioning, production support
 - Labs and Beta Labs support
-

Looking to the Future: What's on our Roadmap

- CX Translation center - drafts, more entry points, contribution suggestions
- Integration of CX and ULS with Wikidata
- On-screen keyboards for input methods in ULS
- Maintain feature parity in PHP and JS i18n
- Review variants support (Serbian, Chinese)
- Roadmap:

[https://www.mediawiki.](https://www.mediawiki.org/wiki/Wikimedia_Language_engineering/Roadmap2014-15)

[org/wiki/Wikimedia_Language_engineering/Roadmap2014-15](https://www.mediawiki.org/wiki/Wikimedia_Language_engineering/Roadmap2014-15)



Thank you for your time!
Language Engineering Team

Wikimedia Foundation

Quarterly Review September 3 2014

