ENGLISH MOTHERFUCKER
Speak my language %app%
или как переводить веб-приложения
DO YOU SPEAK IT?
quickmeme.com

Сергей Фетискин

Берлин

фронтенд-девелопер
10+ лет в разработке
Angular, TypeScript, GraphQL

Twitter: @fetis26

О чем этот доклад

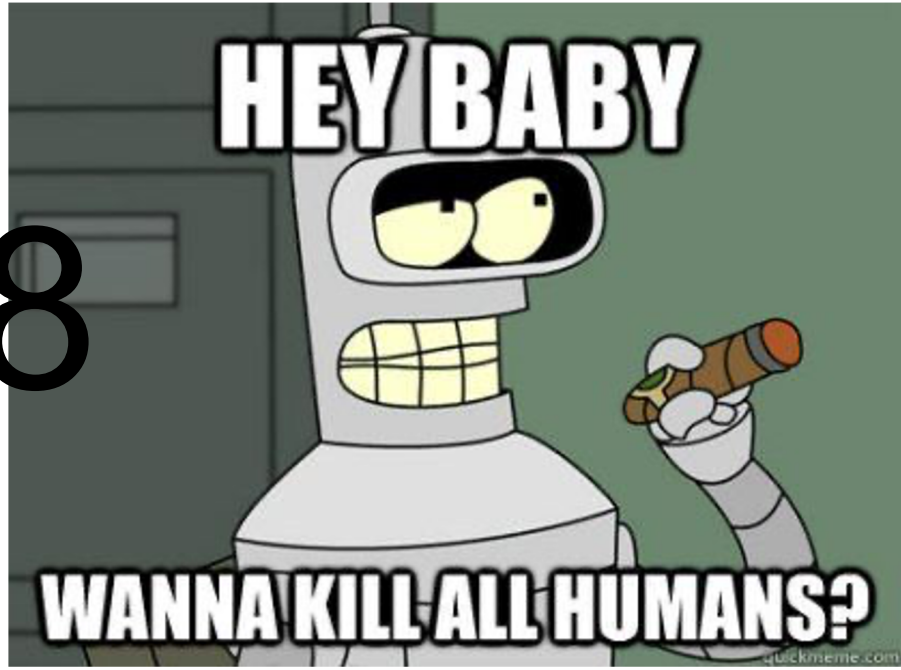
- Проблема
- Термины
- Разработка
- Идеальный процесс
- Форматы
- Платформы для перевода
- BCP 47
- Поддержка в JS и фреймворках

В основном речь пойдет про SPA, хотя многое применимо и для других приложений.

Проблема

Проблема в цифрах

7,8 млрд



общая популяция на Март 2020

Проблема в цифрах

Языков: 90 (10+ млн); 7,8 тыс (ISO 639-3)

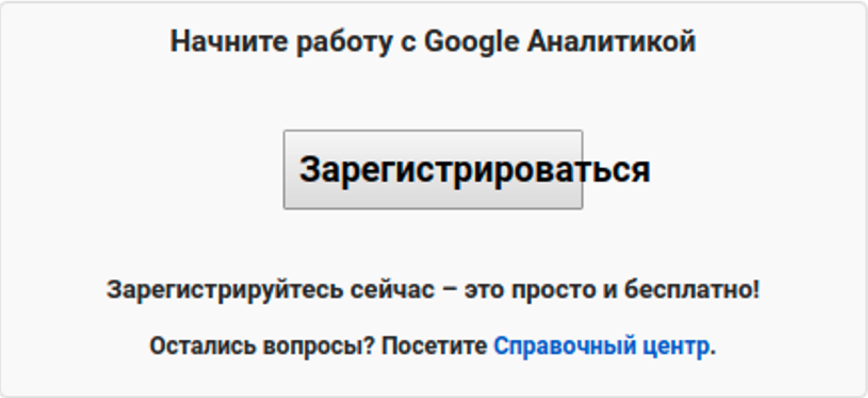
Письменностей: 41

Направлений письма: 3 + комбинации

Временных зон: 38 + летнее время

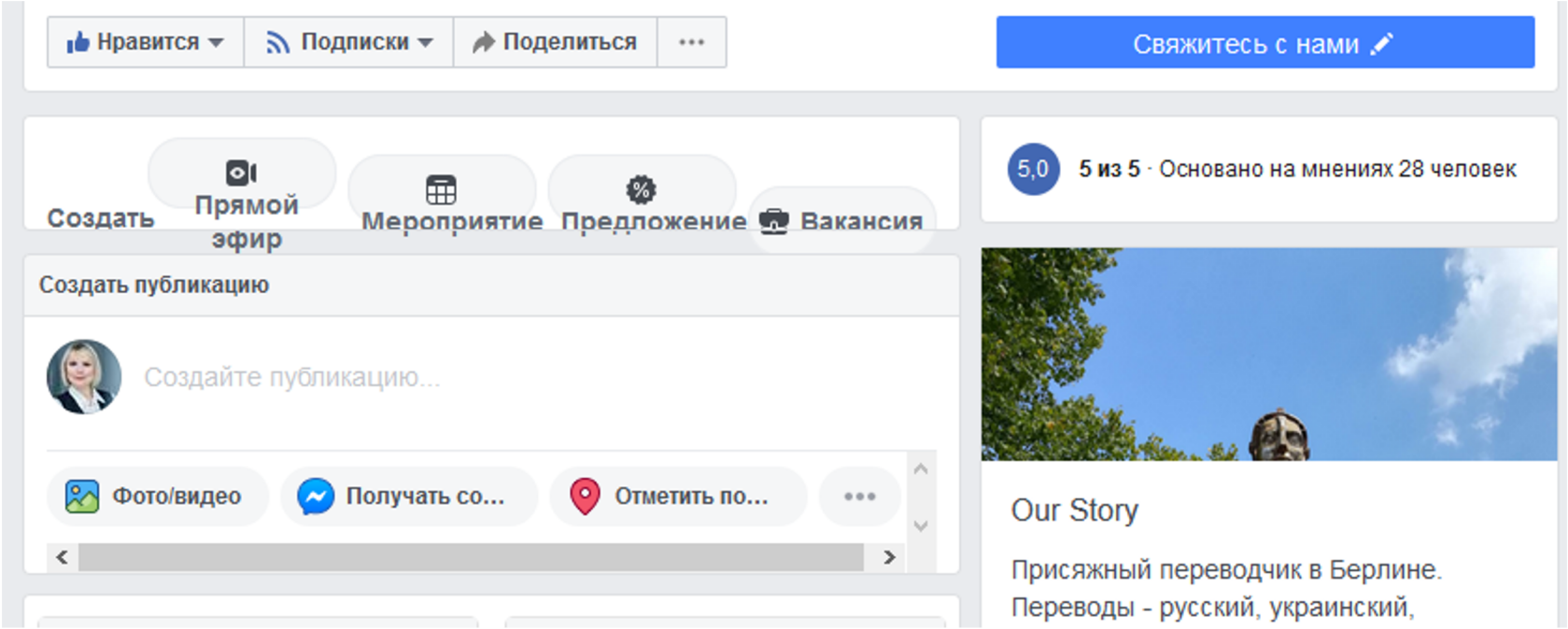
Стран: ~200

Календарей: 8 (григорианский de facto)

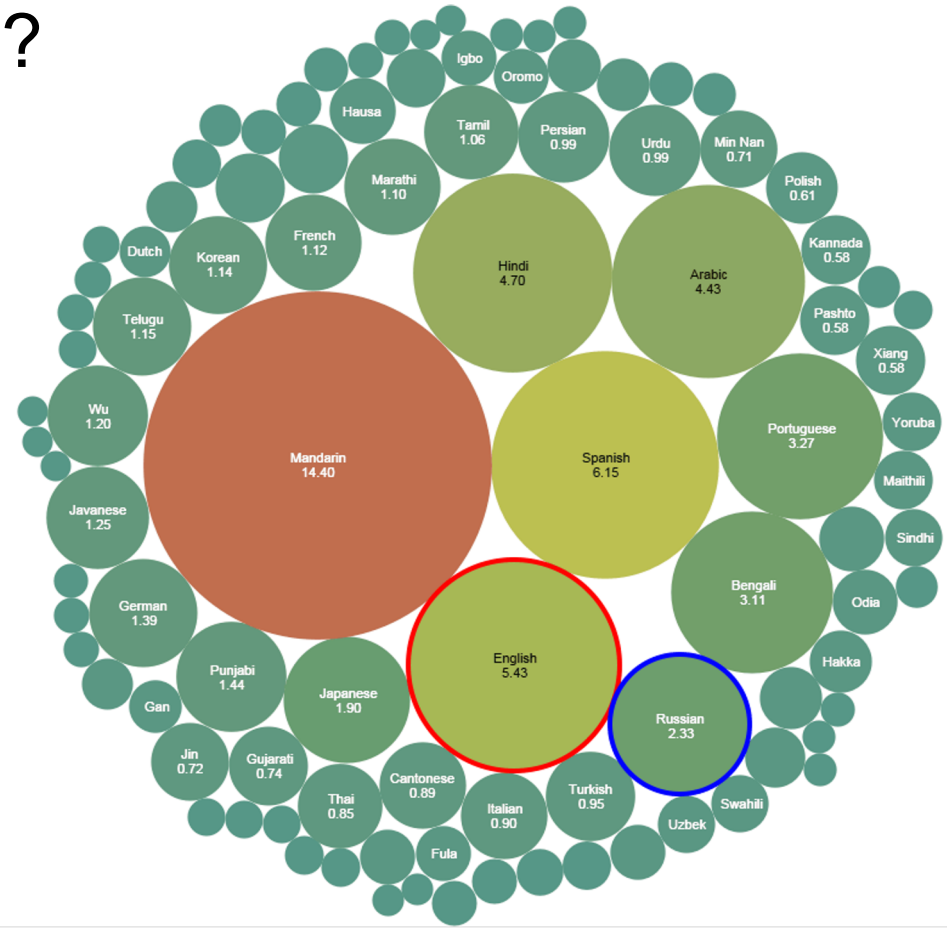


РОЗШИРЕННЯ

In addition to Opera having its own extensions store, the Opera browsers are also compatible with Google Chrome extensions.


Нравится
Подписки
Поделиться
Свяжитесь с нами
Создать
Прямой эфир
Мероприятие
Предложение
Вакансия
Создать публикацию
Создайте публикацию...
Фото/видео
Получать со...
Отметить по...
5,0
5 из 5 · Основано на мнениях 28 человек
Our Story
Присяжный переводчик в Берлине.
Переводы - русский, украинский,

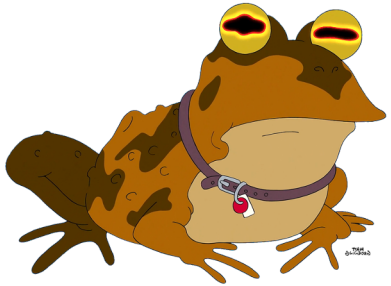
Английский?

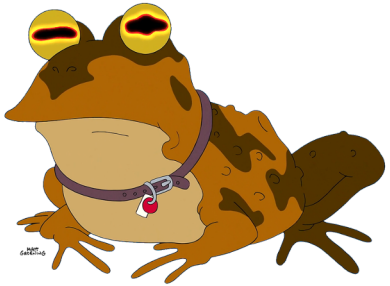

Igbo
Oromo
Hausa
Tamil 1.06
Persian 0.99
Urdu 0.99
Min Nan 0.71
Polish 0.61
Marathi 1.10
French 1.12
Korean 1.14
Dutch
Hindi 4.70
Arabic 4.43
Kannada 0.58
Pashto 0.58
Xiang 0.58
Telugu 1.15
Wu 1.20
Mandarin 14.40
Spanish 6.15
Portuguese 3.27
Yoruba
Maithili
Javanese 1.25
Sindhi
Bengali 3.11
Odia
German 1.39
English 5.43
Hakka
Punjabi 1.44
Japanese 1.90
Russian 2.33
Gan
Jin 0.72
Gujarati 0.74
Cantonese 0.89
Turkish 0.95
Swahili
Thai 0.85
Italian 0.90
Uzbek
Fula

Эсперанто?

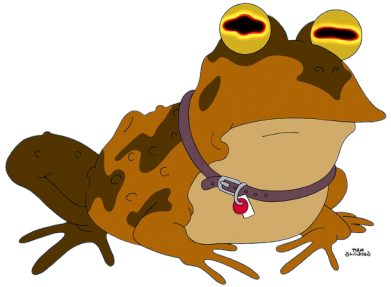
- Не, не слышал.

С момента основания (1887г.) максимум 2 млн пользователей.





Термины



Термины

- Интернационализация (internationalization, i18n, и17я*) — процесс подготовки приложения под разные форматы даты, валюты, ед. измерения, календари, направления текста, набор символов (слава Unicode больше не актуально).

i18n, L10n — это нумеронимы, еще пример a11y

Термины

- Локализация (localization, L10n, л9я*) — адаптация приложения под конкретный регион и язык с использованием средств и17и.

i18n, L10n — это нумеронимы, еще пример a11y

Термины

- Региональные настройки (locale, локаль) — конкретный набор параметров, который задает язык пользователя и параметры отображения данных в интерфейсе.

Примеры: ru, en-US, ru_RU.UTF-8 (POSIX)

Разработка

Проблемы при разработке. Часть 1

- Числительные и мн. число существительных
- Падежи
- Рода
- Форматы даты и времени, разные календари
- Единицы измерения и валюты
- Сравнение строк

Проблемы при разработке. Часть 2

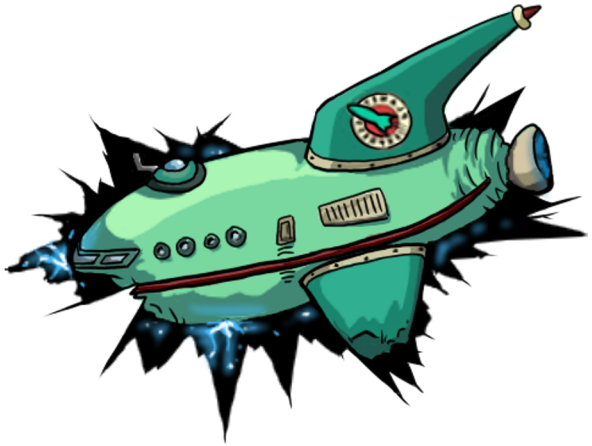
- Разные наборы символов (почти решена в Unicode)
- Разные шрифты
- Разная длина строк в интерфейсе
- *Одинаковый* перевод терминов и названий
- Коллаборация: нетехнические переводчики, утверждение перевода
- Синхронизация исходного кода с переводами

Проблема бизнеса

Перевод не только техническая проблема. Возможно придется перестраивать работу компании и коммуникации региональных департаментов. Нет немецкой версии, нет релиза, например.

Как это обычно происходит

- Excel-документ
- Google Docs
- CSV-файл
- Поддерживаемый вручную JSON



Подытожим

И17я — это сложная техническая задача, которую недооценить на первый взгляд. Мы, как разработчики, хотели бы решать ее быстро и с минимум усилий, используя готовые решения.

International Components of Unicode (ICU)

Международные компоненты Юникод

International Components of Unicode (ICU)

Международные компоненты Unicode — набор C/C++ и Java библиотек для поддержки Unicode, интернализации и глобализации при разработке программ.

Используются в Node.js, Chrome

CLDR (Common Locale Data Repository)

Названия языков, переводы дней недели, месяцев, стандартные форматы даты и валюты.

http://cldr.unicode.org/

https://github.com/unicode-org/cldr

ICU выражения

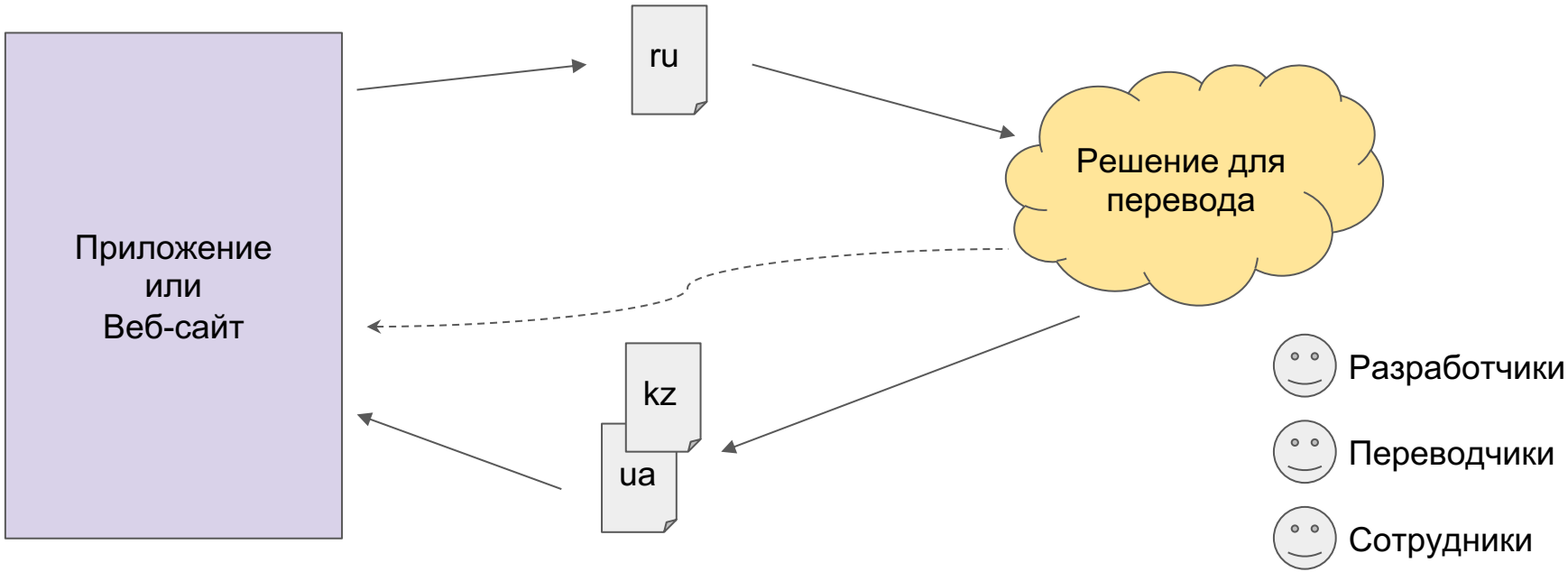
Стандартный способ описания строк, где необходимо сделать подстановку множественного числа или интерполяцию переменных.

Пример ICU выражения

```
<span class="ak-pill" i18n>
    { totalCases, plural,
      =1 {1 case}
      other {{{totalCases}} cases}
    }
</span>
```

Идеальный процесс

Идеальный процесс



Выводы из идеального процесса

1. Универсальный формат для хранения переводов
2. Поддержка автоматического экспорта/импорта строк со стороны фреймворка или библиотеки
3. Решение для перевода в виде облачного сервиса или отдельного приложения. CLI для автоматизации процессов.

Требования для формата хранения

- переносимость
- сохранение контекста
  - номер строки
  - комментарий и подсказки для переводчика
- множественное число
- подстановка переменных

Форматы для хранения

- Excel
- JSON-форматы
- Gettext (PO, POT) +
- XLIFF ++

JSON форматы

- flat JSON
- nested JSON
- JSON with ICU Plurals
- Chrome JSON для расширений ±
- ARB (Application Resource Bundle) ±

```
1.  {
2.    "home.component-based": "Βασισμένο σε στοιχεία",
3.    "home.component-based.p1": "Δημιουργήστε ενσωματωμένα στοιχεία που
    διαχειρίζονται τη δική τους κατάσταση, και στη συνέχεια συνθέστε τα για να
    δημιουργήσετε σύνθετα UI.",
4.    "home.component-based.p2": "Δεδομένου ότι η λογική συνιστωσών είναι
    γραμμένη σε JavaScript αντί για πρότυπα, μπορείτε εύκολα να περάσετε πλούσια
    δεδομένα μέσω της εφαρμογής σας και να κρατήσετε την κατάσταση εκτός του &
    nbsp; DOM.",
5.    "home.declarative": "Δηλωτικό",
6.    "home.declarative.p1": "Το {name} καθιστά ανώφελη τη δημιουργία
    διαδραστικών διεπαφών χρήστη. Σχεδιάστε απλές προβολές για κάθε κράτος στο
    δικό σας εφαρμογή και το React θα ενημερώσει αποτελεσματικά και θα αποδώσει
    τα σωστά στοιχεία όταν τα δεδομένα σας αλλαγές.",
7.    "home.declarative.p2": "Οι δηλωτικές προβολές καθιστούν τον κώδικα πιο
    προβλέψιμο και πιο εύκολο στον εντοπισμό σφαλμάτων.",
8.    "home.welcome": "Καλώς 'Ηρθατε στο {name}!"
9.  }
```

Пример JSON-формата

Что не так с JSON?

Нет единой спецификации.

Каждый придумывает свои велосипеды.

Нет спецификации, соответственно мало инструментов.

Gettext, история вопроса

gettext — система для и17и программ, выпущенная в начале 90-х как GNU gettext. Представляет собой набор утилит для извлечения строк из исходного когда, в переносимый формат и компиляции обратно в бинарный формат.

Портирована на многие платформы, в т.ч. PHP и Node.js

```
#: ad_views.inc:31;127 image/ad_image_views.inc:23 text/ad_text_views.inc:23
msgid "As link"
msgstr ""

#: ad_views.inc:32;128
msgid "Not As link"
msgstr ""

#: ad_views.inc:54
msgid ""
"When using the <em>Advertisement View Type</em> use this field to specify "
"how many ads to display at the same time."
msgstr ""
```

Gettext

+ Устоявшийся формат
+ Много инструментов
+ Интегрируется почти везде

- Меньше возможностей по сравнению с XLIFF
- Чисто текстовый, сложнее преобразовывать

Работа с gettext

1. Разметка строк в исходном с помощью специальной функции gettext

```
/// Comment for translator

printf(_("My name is %s.\n"), my_name);
```

Работа с gettext

2. Извлечение строк в файл шаблона, .pot файл

```
#. Comment for translator

#: src/name.c:36

msgid "My name is %s.\n"

msgstr ""
```

Работа с gettext

3. Перевод, создание локализованной версии,

   .po файл

```
#: src/name.c:36
msgid "My name is %s.\n"
msgstr "Моё имя %s.\n"
```

ru.po (unsaved) — PROJECT VERSION

Upgrade to Pro

Source text — English	Translation — Russian
You have no @status ads.	
Ads cannot be displayed. The ad mo...	
The ad module is <a href="@misconfi...	
Clicked %type ad aid %aid hostid %ho...	Нажато объявление %type %aid %ho...
Ad redirection failed for aid %aid hosti...	Объявление перенаправление не уд...
This advertisement is currently waitin...	Эта реклама в настоящее время жд...
This advertisement has been approve...	Эта реклама была одобрена и в нас...
This advertisement is actively being d...	Эта реклама активно отображается.
This advertisement has been tempora...	Эта реклама была временно отклю...
This advertisement has been unpublis...	Эта реклама не была опубликована...
This advertisement has expired.	

Source text:

Clicked %type ad aid %aid hostid %hostid.

Translation: Needs Work

Нажато объявление %type %aid %hostid.

Translation suggestions:

Нажал %type объявление помощи %aid помощь hostid %hostid.

⌘1 • 100%

PRO Online suggestions are limited to only 10 in the free version. Upgrade to Pro to remove this limitation.
Unlock online suggestions

Comment:

%type and %hostid comment

Edit Comment

Translated: 6 of 526 (1 %) • Remaining: 520

Работа с gettext

4. Компиляция .po файлов в бинарный код, .mo файл

Формат XLIFF

Утвержден организацией OASIS в 2002

Основан на XML

Текущие версии

- 1.2 утверждена в 2008
- 2.1 утверждена в 2018

Формат XLIFF

+ Есть стандарт
+ Много инструментов, есть отдельные редакторы, поддержка во многих платформах для перевода.
- XML, а значит “многословный”

Используется в Angular, iOS и MacOS, Symfony

v1.2 устарела (deprecated), но используется почти везде

Пример XLIFF

```
<app-message-list></app-message-list>

<button i18n="this button makes no action@@btnMakeNice">make it nice</button>

<ak-sidenav #sidenav (openedChange)="openedChanged($event)" class="{{ theme }}">
```

Код на Ангуляре

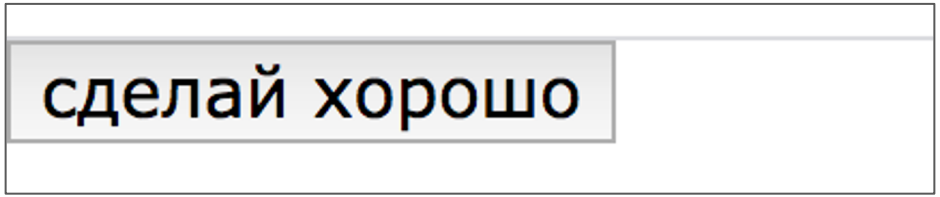
```
<trans-unit id="btnMakeNice" datatype="html">
  <source>make it nice</source>
  <context-group purpose="location">
    <context context-type="sourcefile">app/app.component.html</context>
    <context context-type="linenumber">5</context>
  </context-group>
  <note priority="1" from="description">this button makes no action</note>
</trans-unit>
```

btnMakeNice new ng2.template

this button makes no action

make it nice	сделай хорошо

```
5684        <source>make it nice</source>
5685        <target>сделай хорошо</target>
5686        <context-group purpose="location">
5687          <context context-type="sourcefile">app/app.component.html</context>
5688          <context context-type="linenumber">5</context>
5689        </context-group>
5690        <note priority="1" from="description">this button makes no action</note>
```



Решение для перевода

Как мы выбирали

Обозначили ключевые параметры

Обозначили бюджет: до 50€/месяц

Кандидаты сохранялись в общий документ

Решение для перевода

- Онлайн
- Поддержка нашего формата (XLIFF)
- Поддержка фич XLIFF: примечания, литералы, мн. число
- Поддержка CLI для автоматизации
- Память переводов (translation memory)

All tools should support XLIFF format! Ideally XLIFF 2.0

Applications

OmegaT
https://omegat.org/

POEdit
https://poedit.net

Self-hosted (better not self-hosted)

https://weblate.org/en/
+ paid hosting

https://pootle.translatehouse.org/

https://pontoon.mozilla.org/

Platforms

Crowdin
https://crowdin.com/pricing#annual
CLI support

WebTranslateIt
https://webtranslateit.com/en/plans 50/month
CLI support

Transifex
https://www.transifex.com/pricing/
Probably too expensive, starts from 140/month

PhraseApp
https://phraseapp.com/ $144/month

Lokalise

https://lokalise.co/pricing 40/month
CLI support

~~POEditor~~
~~https://poeditor.com/pricing/ 15$/month~~
Problems with xliff support

MOARR!!

https://www.capterra.com/translation-management-software/
https://blog.gengo.com/16-best-translation-management-systems/

https://opensource.com/article/17/6/open-source-localization-tools

Решения с поддержкой XLIFF

- Lokalise
- Crowdin
- Transifex
- WebTranslateIt
- Phrase
- и прочие

Lokalise

Извлечение строк

Upload

Filename

en.xlf

READY FOR IMPORT | XLF | 310 KEYS 1178 WORDS

Detected language

English (en)

Import the files

or drop / select more files to upload.

Options

- [x] Replace \n with line break
- [ ] Convert placeholders
- [x] Replace modified values
- [x] Tag keys

× new

- [x] Inserted
- [x] Updated
- [ ] Skipped

- [x] Detect ICU plurals
- [ ] Fill empty with keys
- [ ] Hide from contributors
- [ ] Differentiate keys by file
- [x] Cleanup mode

Pre-translate

- [ ] Pre-translate with 100% TM matches

Video tutorial
Uploading files

Lokalise

typeCompany ng2.template property object type %LANG_ISO%.xlf	**English**	company
	Danish	Virksomhed
	French	société
	German	Firma
	Minionish	*Empty*
genderUndefined ng2.template %LANG_ISO%.xlf	**English**	undefined
	Danish	Udefineret
	French	indéfini
	German	nicht definiert
	Minionish	*Empty*
eb8a1b2b302c1762981a2df1913b3c6271f .xlf	**English**	priority
	Danish	prioritet

Lokalise

French	dossiers 1
German	Fälle 1

Insert source ⌥0 | 72 Tab - next | Mark as reviewed ⌘E

100% Fälle <x id="INTERPOLATION" equiv-text="{{caseList?.itemsTotal || 0}}"/> ⌘⌥1
results <x id="INTERPOLATION" equiv-text="{{caseList?.itemsTotal || 0}}"/>

91% Fälle <x id="INTERPOLATION" equiv-text="{{ caseList?.itemsTotal }}"/> ⌘⌥2
cases <x id="INTERPOLATION" equiv-text="{{ caseList?.itemsTotal }}"/>

81% Reparaturfälle <x id="INTERPOLATION" equiv-text="{{repairCaseList?.itemsTotal || 0}}"/> ⌘⌥3
repair <x id="INTERPOLATION" equiv-text="{{repairCaseList?.itemsTotal || 0}}"/>

78% Administrative <x id="INTERPOLATION" equiv-text="{{adminCaseList?.itemsTotal || 0}}"/> ⌘⌥4
administrative <x id="INTERPOLATION" equiv-text="{{adminCaseList?.itemsTotal || 0}}"/>

Ergebnisse <x id="INTERPOLATION" equiv-text="{{caseList?.itemsTotal || 0}}"/> ⌥1

Lokalise

Translation memory ⌘G

send

Results: 10

All target languages

Exact match

Search target

Include regions

envoyer

send

3 Sep, 2019 in case management frontend

absenden

send

3 Sep, 2019 in case management frontend

send

send

17 Feb, 2020 in case management frontend

Lokalise

Lokalise

```
case-management-frontend git:(master) ✗  ➜
```

Загрузка свежего перевода

Языковой тег BCP 47

BCP47

Языковой тег (language tag) для обозначения языка.

Разбивается на подстандарты, которые описывают отдельные части.

Используется в:

- HTTP-заголовки
- атрибут lang
- Internationalization API (Intl.xxx)

BCP 47 — Best Current Practices 47 от организации IETF

Структруа BCP47

language-extlang-script-region-variant-extension-privateuse

`ru` - русский

`Cyrl` - кириллица

`RU` - Россия

`petr1708` - Орфография согласно Петровской реформе 1708

IANA registry

Используйте его для обозначения языка

Неправильно: ru_RU. Унаследвано от POSIX стандарта

Правильно: ru, en, de

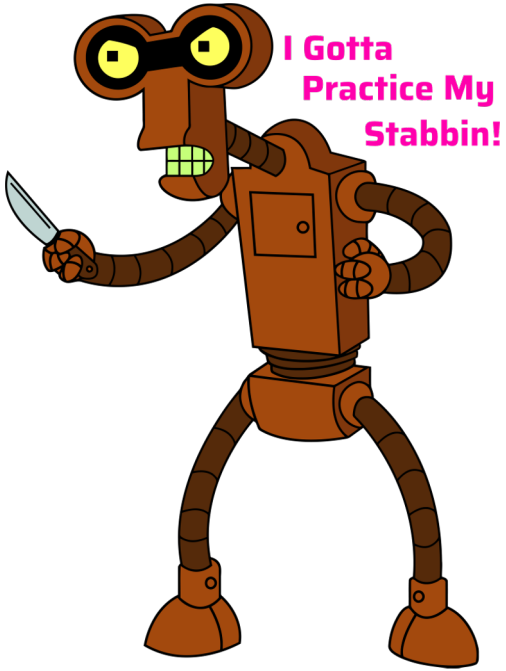
en, en-GB, fr, fr-CA, pt, pt-BR когда есть **два варианта** одного языка

Интересные примеры

ru-Latin — транслит

tlh — клингонский язык

Прочие вопросы



Когда делать и17ю?

В начале разработки

- процесс более гладок
- возможные проблемы сразу видны
- больший процент лишних строк (интерфейс еще не стабилен)

В конце разработки

- **тяжело** сразу добавить для всего проекта
- меньший процент лишних строк (интерфейс стабилен)

Всегда отталкивайтесь от бизнес-задач

Выбор основного языка

- Язык команды (обычно английский)
- Если уже есть готовый продукт, то оставьте как есть
- Использовать id, а не текст

Флаги или названия?

Флаги

- проблемы с доступностью, потому что это картинка
- достаточно мелкие; если их много рябит в глазах
- многие флаги похожи (Бельгия и Германия, Россия и Голландия)
- неоднозначно для некоторых стран Канада (2), Швейцария (4), Бельгия (3), Индия (2), Россия (сюрприз, однако много регионов имеют 2 и более официальных языка). Таких стран много.
- не характеризуют язык (какой флаг для английского, например?)

Флаги или названия?

Страны или коды языка

- простой текст, нет проблем с доступностью
- может быть записан на родном языке
- более однозначно определяет язык
- поддерживающая иконка желательна

JS и фреймворки

Поддержка во фреймворках

Angular

- встроенный i18n атрибут *
- ~~ngx-translate *~~
- locl
- angular-l10n
- transloco

React

- react-i18next
- react-intl *
- lingui.js *

Vue.js

- vue-i18n

* извлечение строк через CLI

JS библиотеки

fluent.js

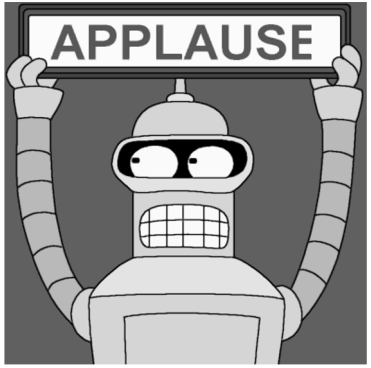
i18next.js

globalize.js

format.js

lingui.js

Сергей Фетискин

Берлин

Twitter: @fetis26