



# Локализация веб-проектов

Олег Алистратов

# I18N и L10N

Интернационализация и локализация

# Локализация

## Адаптация интерфейса:

- Тексты, сортировка и сравнение
- Форматы чисел, дат и валют
- Символы и типографика
- Пиктограммы, рисунки и цвета
- Средства ввода (клавиатура)

# Локализация

## Адаптация содержания и поведения:

- Способ ведения бизнеса  
(например, бухгалтерский учет)
- Культурозависимые объекты и идеи
- Географическое положение
- Юридические требования

# Интернационализация

I18N — проектирование и разработка с целью легкой локализации:

- Убираем ограничения (например, внедряем Unicode)
- Проверяем поддержку функций I10n заранее (например, двунаправленный или вертикальный текст)
- Включаем «локаль» (системные функции форматирования)
- Разделяем код и локализуемые ресурсы

# Объекты локализации

Я

# Мега-сайт

- Новости
- Товары
- О компании

# Мега-сайт локализованный

- Новости
- Товары
- О компании
- News
- Products
- About the company



## Мега-сайт – 2

- Новости
  - Товары
    - Молотки
    - Трубы
    - Напильники
  - О компании
- News
  - **Products**
    - **Hammers**
    - **Tubing**
    - **Files**
  - About the company

## Мега-сайт – 3

- Новости
  - Товары
    - Молотки
    - Трубы
      - Труба 20 мм
    - Напильники
  - О компании
- News
  - Products
    - Hammers
    - Tubing
      - Tubing 20 mm
      - Tubing 1.5"
  - About the company

# Мега-сайт – 4

- Доставка
  - Курьером
  - Почтой
- Оплата
  - Наличными
  - Наложным платежом
- Delivery
  - Priority overnight
  - Saturday delivery
  - Washington, DC delivery
- Payment info
  - Visa
  - Mastercard
  - C.O.D.

# Взболтать, но не смешивать!

- Язык → Интерфейс
- Регион → Поведение

# Определение локали

Я

# Accept-Language

- Accept-Language:  
uk, ru;q=0.9, en-GB;q=0.8, en;q=0.7
- Accept-Language:  
de-DE, de-CH, de-AT
- Accept-Language: de

# GeoIP

- Страна ≠ Язык
- Невероятные приключения итальянцев в России

# Доменные имена

- en.example.com
- ru.example.com
- uk.example.com
  
- ru.user.hosting.com



## Доменные имена - 2

- [www.example.co.uk](http://www.example.co.uk)
- [www.example.ru](http://www.example.ru)
- [www.example.com.ua](http://www.example.com.ua)
- [www.example.com](http://www.example.com)
- Multi-domain Authentication (MDA)

## Доменные имена - 3

- [www.example.com/en/](http://www.example.com/en/)
- [www.example.com/ru/](http://www.example.com/ru/)
- [www.example.com/uk/](http://www.example.com/uk/)

# Определили – запоминаем

- Язык
- Регион
- Валюту
- Форматы дат и т.п.

...и разрешаем изменить по вкусу!

А еще: в узбекском – шрифт,  
в венгерском – порядок сортировки...

# Технологии

Я

# Все уже придумано до нас

- GNU gettext
- POSIX locale

# Вокруг gettext'a

Файлы ресурсов:

- ru/LC\_MESSAGES/animals.mo

bird → ПТИЦА

cat → КОТ

- uk/LC\_MESSAGES/animals.mo

bird → ПТИЦЯ

cat → КІТ

Я

# Вокруг gettext'a

## Использование:

- ```
textdomain("animals");  
print(gettext("bird"));
```
- ```
print(gettext(  
"Little bird! O little bird!\n"  
"I wonder at what thou doest,\n"  
"Thou singing merry far from me,\n"  
"I in sadness all alone!"  
));
```

# Искусственные ключи

Использование:

- `print(gettext("msg_bird_name"));`
- `print(gettext("msg_bird_poem"));`



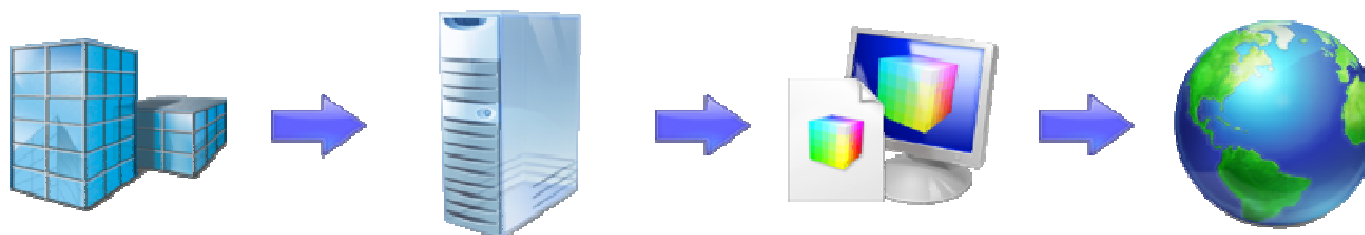
# Искусственные ключи

- Не засоряются исходники
- Решается проблема контекста
- Трудно отслеживать изменения
- Без перевода нет осмысленного текста

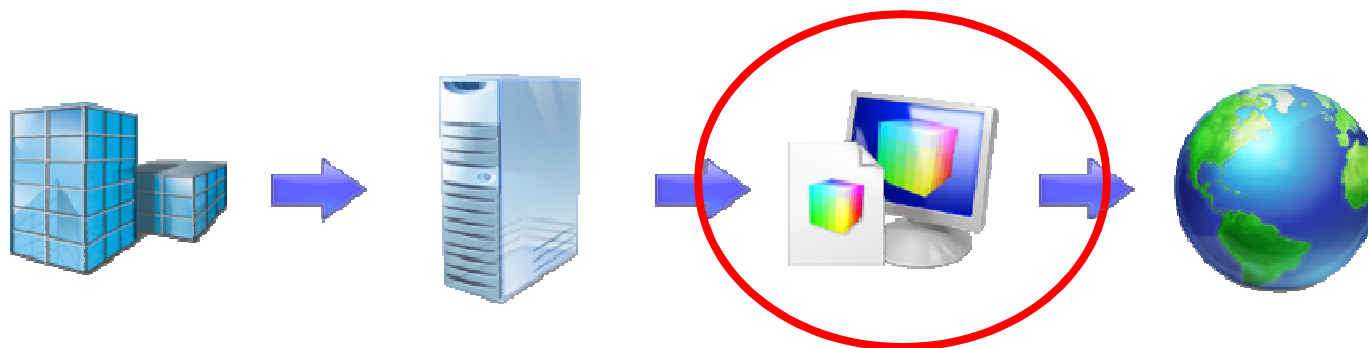
# Усложняем

- "logo" = object {  
    src = "logo.jpg",  
    width = "120",  
    height = "80",  
    title = "ABC Ltd."  
}
- "logo" = object {  
    src = "logo.uk.jpg",  
    width = "200",  
    height = "90",  
    title = "ТОВ «ABC»"  
}

# Точка наложения перевода



# Точка наложения перевода



# Подготовка исходных текстов

# Хороший стиль

- Баннер: код
- Код баннера для вставки на страницу

# Полные фразы

- printf (  
    "Файл защищен от %s",  
    rw ? "записи" : "чтения"  
)
- printf (  
    rw ?  
    "Файл защищен от записи" :  
    "Файл защищен от чтения"  
)

# Форматированные строки вместо конкатенации

- "в городе: " + "Киев"
- printf("в городе: %s", "Киев")



## Избегаем специальных символов

- "Иван&nbsp;Петрович"
- "Иван Петрович"

# Кому специальный, а кому и родной

- 'Кам'янець-Подільський'
- 'Кам\`янець-Подільський'

# Верстка

Я

# Проверяем длинными фразами

|         |        |            |
|---------|--------|------------|
| Новости | Товары | О компании |
|---------|--------|------------|

|      |          |                   |
|------|----------|-------------------|
| News | Products | About the company |
|------|----------|-------------------|

|      |          |                              |
|------|----------|------------------------------|
| News | Produits | À propos<br>de cette société |
|------|----------|------------------------------|

# Используем CSS для заглавных и строчных символов

- `<p>Сегодня понедельник</p>`  
`<p>Понедельник, 20 сентября</p>`
- `<style>`  
`.ucfirst:first-letter {text-transform: uppercase}`  
`</style>`  
  
`<p>Сегодня понедельник</p>`  
`<p class="ucfirst">понедельник,`  
`20 сентября</p>`

## Атрибуты *lang* и *dir*

lang="en" dir="ltr"

lang="ar" dir="rtl"

- Localization

- الاقلمه والتدوي

# Избегаем текстов в графике

- Рисованные заголовки – зло
- Скриншоты, например, тоже зло
- Если очень хочется – можно генерировать программно:  
``

# Поменьше разметки

- **gettext(**  
**<a href="http://example.com">**  
**Example**  
**</a>**  
**)**
- **<a href=**  
**"gettext('http://example.com')">**  
**gettext('Example')**  
**</a>**



# Формы множественного числа

Я

# Правильная форма

- Found 42 file(s)
- Found 1 file  
Found 42 files
- Найден 101 файл  
Найдено 42 файла  
Найдено 15 файлов

# Ищем решение

- `printf("Found %d file%s", n,  
n == 1 ? "" : "s");`
- `if (n == 1)  
 printf ("Found %d file", n);  
else  
 printf ("Found %d files", n);`
- `int idx = plural_form(n);  
printf(strings[FOUND_MESSAGE][idx], n);`

# Список форм в файле ресурсов

- lang="en"  
Две формы

[0] Found %d file  
[1] Found %d files

- lang="ru"  
Три формы

[0] Найден %d файл  
[1] Найдено %d файла  
[2] Найдено %d файлов

## Пример: английский язык

- ```
function plural_form(int n) {  
    return n != 1 ? 1 : 0;  
}
```
- ...а также немецкий, норвежский, греческий, итальянский, эстонский, эсперанто и многие другие языки

## Пример: русский язык

- ```
function plural_form(int n) {  
    return  
    n % 10 == 1 && n % 100 != 11 ?  
    0 : n % 10 >= 2 && n % 10 <= 4 &&  
    (n % 100 < 10 || n % 100 >= 20) ?  
    1 : 2;  
}
```
- ...а также украинский, белорусский, сербский и хорватский

## Пример: китайский язык

- ```
function plural_form(int n) {  
    return 0;  
}
```

## Усложняем задачу

- Found 10 files in 5 directories
- Found %d %s in %d %s
- Found {NUM\_FILES} {FORM\_FILES}  
in {NUM\_DIRS} {FORM\_DIRS}



# Кодировки

Я

# Unicode

## Преимущества UTF-8:

- Цифры, знаки препинания и латиница отображаются всегда
- Экономим объем (трафик)
- Нет путаницы с *little endian* и *big endian*
- В UTF-16 не всегда два байта (только символы UCS-2)
- Поддержка редакторами, БД, XML-процессорами...

## 8 бит

- Несколько языков могут поместиться в одной 8-битной кодировке:  
английский, русский, украинский — Win-1251
- А могут и не поместиться:  
русский и немецкий

# Даты, валюты, единицы измерения

# POSIX

- `setlocale(LC_ALL, "uk_UA.UTF-8");`
- `sprintf()`, `sscanf()`
- `strftime()`, `strptime()`
- `strfmon()`

# Локаль можно обработать напильником

- \$ LANG=uk\_UA date +"%e %B"  
10 Квітень
- \$ LANG=uk\_UA date +"%e %B"  
10 квітня

# Строги при выводе, лояльны при вводе

- out: `printf()`  
21,00 грн.  
21.00 €
- in: `scanf()`  
21.00  
21,00  
21-00

# Номера телефонов

- (044) 200-10-20
- +380 44 200-1020  
8 (044) 200-10-20  
200-10-20



# Локальное время

- Мы работаем с 9<sup>00</sup> до 18<sup>00</sup>  
(MSK, UTC+3)
- Мы працюємо з 8<sup>00</sup> до 17<sup>00</sup>  
(EET, UTC+2)

# Anno Domini

- 6244 – Древнеегипетский
- **5768 – Еврейский**
- 5208 – Древнеиндийский
- 5127 – Календарь Майя
- **4706 – Китайский**
- 2761 – Древнеримский
- 2757 – Вавилонский
- **2552 – Буддистский**
- **2008 – Григорианский**
- **1429 – Мусульманский**
- **1387 – Персидский**
- 216 – Календарь Французской революции
- 164 – Бахаи

# Единицы измерения

- Длина: 100 м
- Length: 328.08399 feet
- Length  $\approx$  328 feet

Желаю успеха!  
Бажаю успіху!  
Je vous souhaite du succes!  
Le deseo éxito!  
Ich wünsche ihnen Gelingen!  
Сизга мен мувафакият тилайман!

**Я**ndex