

Интернационализация и Локализация. Введение.



Основные понятия:

Глобализация

подготовка продукта к работе на разных рынках. Сюда относятся различные аспекты маркетинга, дизайна и технологической реализации.

Интернационализация

(сокращенно — i18n) — обеспечение максимальной универсальности продукта, создание прочного фундамента для локализации и перевода.

Локализация

(сокращенно — l10n) — адаптация продукта к определенным региону, культуре и языку.

Перевод

(сокращенно — t9n) — собственно перевод продукта на новый язык.



Тестирование интернационализации и локализации часто связана с тем, что эти процессы воспринимаются лишь как перевод пользовательского интерфейса и документации, тогда как на самом деле это куда более сложные задачи.



Интернационализация

Интернационализация — процесс разработки приложения, при котором код самого приложения независим от любых языковых и культурных особенностей региона/страны. Суть интернационализации в том, чтобы сделать процесс локализации проще, дешевле и быстрее. Реализацию интернационализации обычно начинают на ранних этапах проекта, чтобы подготовить продукт к будущей локализации. Во время процесса интернационализации определяют, что будет изменяться для будущих локалей (например текст, изображения и т.п.) и выносят эти данные во внешние файлы. Также во время интернационализации нужно добавить возможность изменять календари, форматы даты, времени, цифр, денежных символов и в целом символов, специфичных для определенных языков и многое другое. Как итог, в идеальном варианте, добавление новой локали не должно требовать от нас изменения исходного кода продукта!

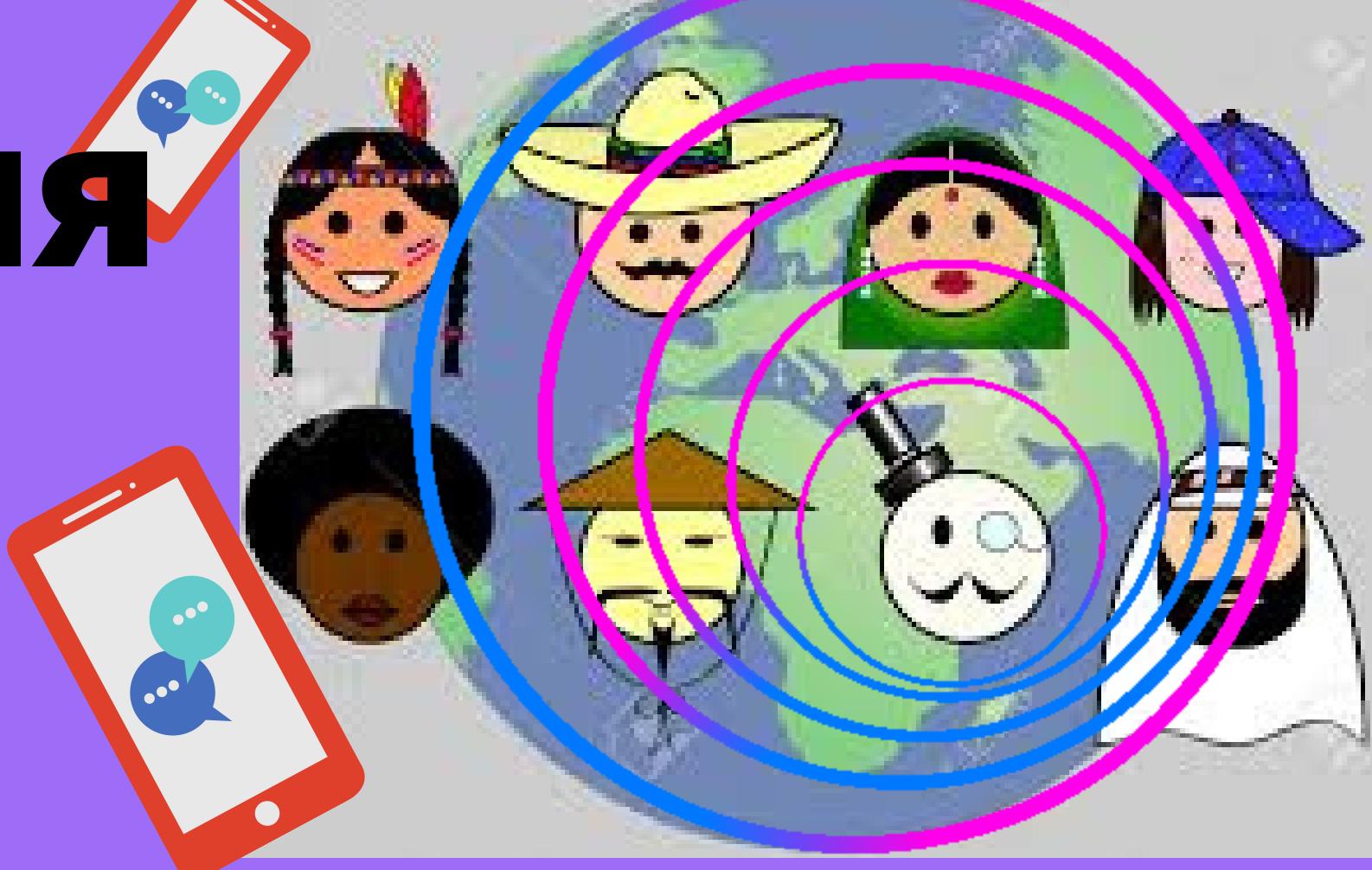
Чек-лист Интернационализации

- Установка и удаление продукта при использовании локали. Раз мы можем и должны использовать различные локали, то мы должны иметь возможность устанавливать и удалять приложение, используя любой язык.
- Настройка приложения на локали, отличной от дефолтной. Все изменения в настройках должны сохраняться и использоваться.
- Приложение должно работать на окружении, локаль которого отличается от дефолтной локали нашего приложения.
- При запуске приложения на новой локали вам нужно проверить, что форматы даты, времени, спец.символы алфавита, денежных валют, календари и прочие вещи, связанные с культурными особенностями можно переключать в приложении без каких-либо ошибок (вручную или автоматически). Как правило такие ошибки выявляются при первом запуске новой локали.
- Локализованный текст, изображения и прочие ресурсы не вшиты в приложение. Баги такого рода найдутся достаточно быстро, ведь, к примеру, оставшиеся английские слова будут сразу видны на фоне азиатского текста.
- Также обратить внимание на то, что документы (гайды и т.д.) переключаются на нужный вам язык при смене локали.

Локализация

Зачем же нам нужна локализация? В первую очередь — распространение нашего продукта на глобальный рынок. Чем в большем количестве стран он будет доступен, тем больше ширится список потенциальных и фактических клиентов. Ведь всем приятно пользоваться продуктом на родном языке, даже если вы хорошо владеете иностранными. Как следствие — доход и прибыль увеличиваются! Стоит сразу сказать, что локализация != перевод.

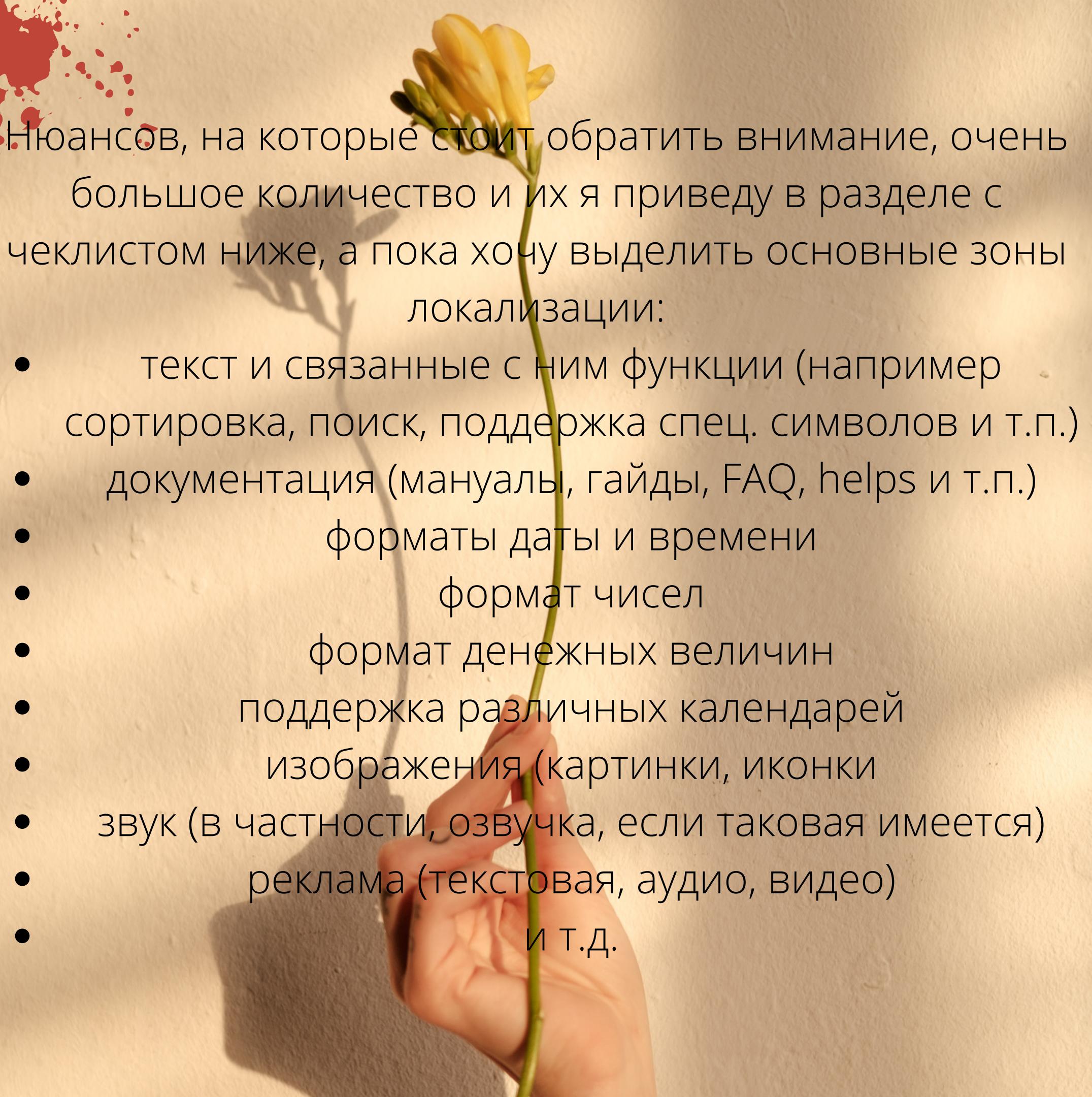
Локализация — это перевод и культурная адаптация продукта к особенностям определенной страны или региона.



Это должен быть продукт, в котором региональные и культурные особенности (текст, картинки, форматы даты, времени и т.п.) будут вынесены в отдельные блоки, которые будут подгружаться при использовании того или иного региона/страны. Данный набор ресурсов называют "локалью" (locale).



- Помогает избежать хардкода для ресурсов, которые должны быть переведены. Если вы переключились на какой-то регион/язык, но видите текст не на ожидаемом языке — это явно баг, если вы изначально не решили, что определённые места не будут переведены (подробнее об этом можно узнать в разделе о псевдо-локализации ниже)
- Файлы, нужные для локали, уже предопределены в требованиях.
- В случае, если нужно добавить новую локализацию, то всё, что нужно сделать — создать нужные данные для вашего региона/страны и подгрузить их в приложении.
- Баги, найденные при тестировании локализации, как правило, будут связаны непосредственно с локалью, а не с самим приложением. Если баг связан с кодом приложения, то скорее всего проблема связана с тем, что данный ресурс не был вынесен в локаль или не был переведён, т.е. это баг интернационализации вашего приложения, либо самой новой локали.



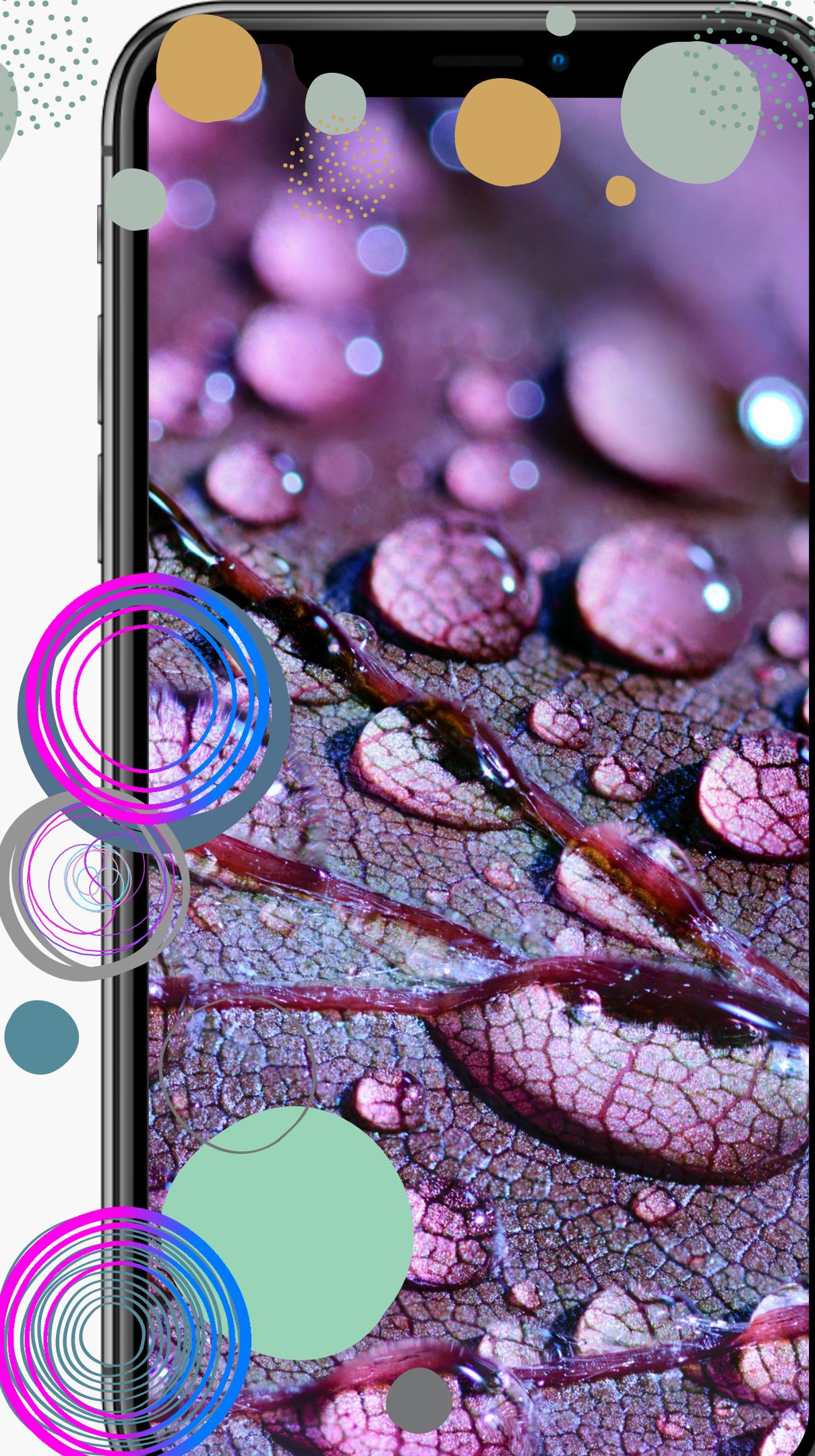
Нюансов, на которые стоит обратить внимание, очень большое количество и их я приведу в разделе с чеклистом ниже, а пока хочу выделить основные зоны локализации:

- текст и связанные с ним функции (например сортировка, поиск, поддержка спец. символов и т.п.)
- документация (мануалы, гайды, FAQ, helps и т.п.)
- форматы даты и времени
- формат чисел
- формат денежных величин
- поддержка различных календарей
- изображения (картинки, иконки)
- звук (в частности, озвучка, если таковая имеется)
- реклама (текстовая, аудио, видео)
- и т.д.

Тестирование локализации

Метод псевдо-локализации:

При тестировании локализации можно выделить 2 подхода: когда у нас есть локаль и когда её нету. С первым вариантом всё просто, берём локаль и проводим необходимый комплекс операций для проверки её качества. Но, как ни странно, 2й вариант тоже имеет место на жизнь и будет крайне полезен, когда интернационализация продукта только завершилась и у вас ещё нету новой локали. В данном случае нам очень поможет техника псевдо-локализации. В рамках этого подхода вам нужно подготовить псевдо-локаль (сделать это может и сам тестировщик, не имеющий знаний нужного языка), используя любой другой язык и добавить в файлы необходимые вам данные, например спец. символы определенного языка. Также можно добавить большее количество символов в строки, чтобы проверить будет ли обрезаться текст, ведь текст может стать длиннее после перевода на новую локаль и не вместиться в отведённое для него пространство. Ещё такой подход поможет вам найти проблему с объединением строк и проблемы, связанные с отображением шрифта, если такие имеются. Для того, чтобы такие проблемы можно было проще найти, то в начало и конец строки можно добавить какой-нибудь символ.



Чек-лист Локализации: Региональные и культурные особенности

Календари

Одно из основных различий между календарями заключается в том, что каждый из них может иметь разное значение года для текущей даты, в зависимости от старта исчисления времени текущего периода.

Продолжительность года и месяцев может варьироваться, также как и способы обработки високосных лет. Даже первый день недели может начинаться в день, отличный от понедельника (Европейские страны) или воскресенья (США). Ну а напоследок стоит отметить, что в некоторых локалях может использоваться более одного календаря

Бывают:
Григорианский,
японский, буддийский,
хиджринский,
еврейский лунный,
тайваньский

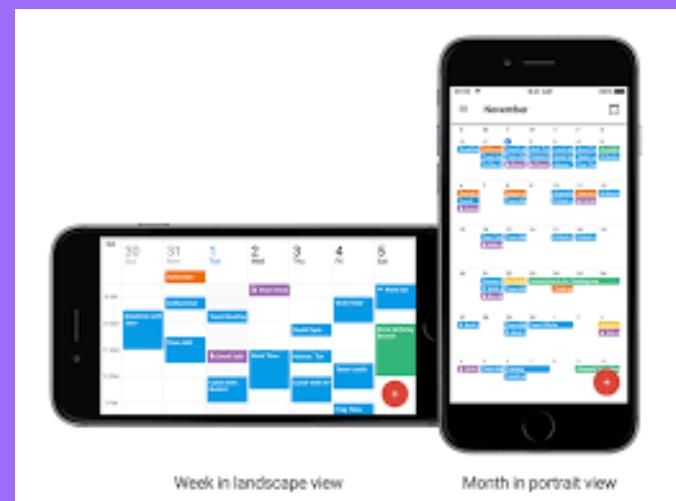
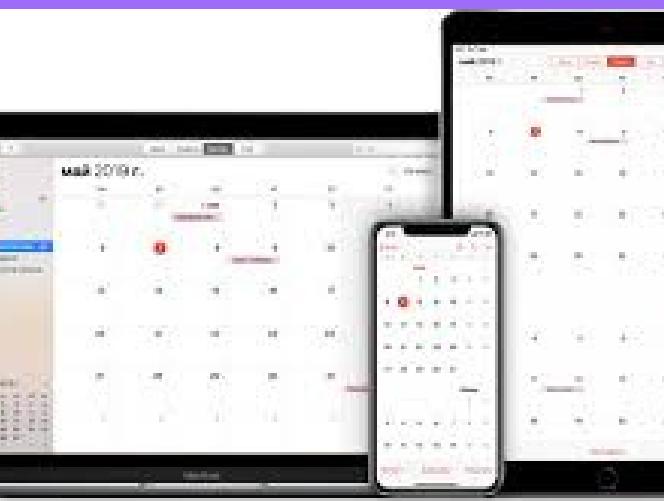
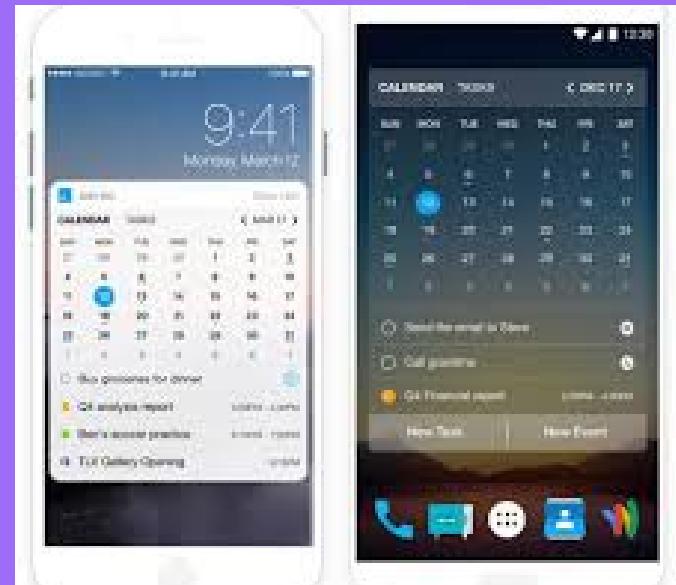




Форматирование дат

Дата подразумевает такие составляющие как день, месяц и год. Но в мире не существует какого-либо общего используемого стандарта отображения этих данных, а в некоторых случаях различия могут существовать даже в разных регионах одной страны.

Для отображения даты обычно используется маска вида DD/MM/YYYY, где D — день, M — месяц, а Y — год. При этом количество букв в каждой части обозначает количество символов, которые будут указаны в каждой секции. Например для маски вида DD/MM/YYYY дата будет выглядеть как 21/11/2021, а для маски вида DD/MM/YY — 21/11/21.



По полноте отображения формата даты:

короткий (short) — 21.11.2021

средний (medium) — 21 Jun 2021

длинный (long) — 21 June 2021

полный (full) — Monday, 21 June 2021

Существуют и символы-разделители.

В зависимости от формата, разделители могут отличаться. Самые распространенные — косая черта (/), точка (.), пробел ().

Каждый отдельный элемент в дате может быть записан в своём формате, который может меняться в зависимости от региона.

Необходимо учитывать различные элементы и составляющие даты, которые могут быть использованы:

День недели (Sunday)

Месяц (November)

День месяца (21)

Год (2021)

Эра (AD)

День года (325)

Неделя года (46)

Неделя месяца (3)

День недели в месяце (7)

Квартал (4)

Первый день недели

Как упоминалось выше, первый день недели может варьироваться в зависимости от региона/страны. Этот пункт стоит выделить отдельно, т.к. часто он остаётся незамеченным, но от этого он не становится менее важным.

Существуют специфическое написание дат для разных регионов/стран. Для примера даты, написанные в формате таких стран как США, Мексика, Япония (используется длинный формат даты — long date):

USA -Tuesday, October 12, 1954

Mexico: martes 10 de octubre de 1989

Japan: 1989年11月05

Названия месяцев и дней недели в разных регионах различаются.

Однако есть и другие заметные отличия. В Мексике день стоит перед месяцем, всё пишется строчными буквами и добавлен артикль «de». В Японии день недели не показан, а для отображения слов "год, месяц, день" используются иероглифы.

Сегодня существует достаточноное количество различных API, какой стране вы находитесь, которые прекрасно работают с датами и примеры выше показывают, почему стоит их использовать. Они не только будут обрабатывать формат правильно, но также будут отображать правильные переводы для длинных дат, что позволит сократить расходы на перевод.

Пример с коротким форматом даты:

United States: 10/12/54

Mexico: 12/10/54

Japan: 54/10/12

Например:

В зависимости от того, в

дата, указанная как 07/04/01,

может означать:

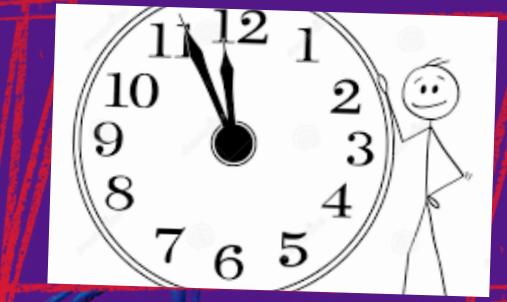
United States: July 4th, 2001

Mexico: April 7th, 2001

Japan: April 1st, 2007

Форматирование времени

Самым распространённый пример - это "часы, минуты, секунды", формат их написания и разделительные символы могут сильно различаться. Различия могут быть даже в рамках разных регионов одной страны.





Примеры возможных вариантов записи времени

По полноте отображения даты можно выделить основные форматы:

короткий (short) — 23:20 (часы, минуты)

средний (medium) — 23:20:59 (часы, минуты, секунды)

длинный (long) — 23:20:59 GMT+3 (часы, минуты, секунды, часовой пояс)

полный (full) — 23:20:59

Kiev Standard Time (часы, минуты, секунды, полное локальное обозначение часового пояса)

Различные варианты отображения форматов времени:

Использование 12-часового или 24-часового формата;

Разделительный символ для часов, минут и секунд:

- самый распространённый (:);
- в некоторых регионах "h", "m" и "s".

Отображение часовых поясов:

Наиболее распространенным форматом отображения часового пояса является формат вида: GMT +3:30.

При тестировании часовых поясов вы должны принять во внимание следующее:

Не все страны используют местные
названия для часовых поясов.

Аббревиатуры часовых поясов не
的独特。

Не во всех странах используется летнее
время, при этом летнее время может
начинаться и заканчиваться в разные
дни для разных стран.

Один часовой пояс может иметь
много разных названий в
зависимости от страны и языка.
Смещение может быть не только
на пол часа или час, но также и
на 15-минутное значение

Форматирование чисел

Крайне важно обращать внимание на форматы, особенно если вы работаете с проектом в финансовой сфере, ведь любое изменение записи формата числа может непредсказуемо повлиять на результаты финансовых операций в вашем приложении:

1. Разделитесь для тысяч (,) (.)();
2. Разделитель дробной части (.) (,);
3. Отображение отрицательных чисел:

-123 123- (123) [123];

5. Группировка чисел
6. Местоположение символа процента (%)
7. Нумерованные списки

4. Отличие чисел от их десятичного аналога
например:

Латинские числа 0 1 2 3 4
Арабские числа ٠ ١ ٢ ٣ ٤ ٥ ٦
Японские числа 〇 一 二 三
Корейские числа 〇 일 이 삼

Форматирование значений валюты

В качестве символа валюты будет использоваться заранее определенный символ (например европейский евро — €, Доллар США — \$, и т.д.).

Местоположение символа валюты может быть как до, так и после цифрового значения суммы.

Значение температуры

Как правило используются значения в Цельсиях (3°C) и Фаренгейтах (3°F).

Системы мер

- Метрическая система мер (метры, литры, граммы и т.д.);
- Английская система (имперская система)- это (футы, дюймы, фунты и т.д.).

Буквенные и числовые значения

Алфавит/ Шрифт

Правильное отображение шрифтов и связанные с этим нюансы. Здесь очень много нюансов. Например:

1. Усечение/обрезка слов и предложений.
2. Не вместимость текста в нужный элемент из-за его увеличения после перевода.
3. Закодированный текст.
4. (Не) отображение спец. символов для данного языка.
5. (Не) отображение конкретного шрифта.
6. Объединение нескольких строк в одну.
7. Размер символов шрифта на экране (к примеру некоторые иероглифы могут быть слишком маленькие на вашем элементе, из-за чего их крайне сложно или невозможно прочитать).
8. Учитывать возможность наличия двух государственных языков.

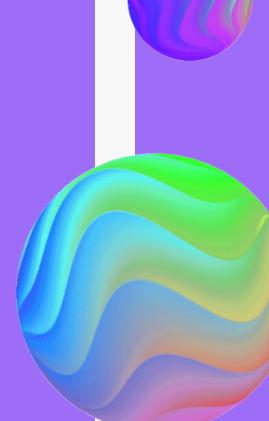
Ориентация текста

Текст в разных регионах может читаться по-разному: у нас принято читать текст слева-направо сверху-вниз (построчно), в арабских странах, например, читают справа-налево (построчно), а в азиатских странах есть следующие вариации: чаще всего встречается вариант справа-налево и сверху-вниз (в столбик), реже — построчно.



Сортировка строк

Каждый язык имеет свой уникальный набор правил (иногда более одного) описывающий как языковые строки должны быть «отсортированы» в упорядоченный список. Ожидания пользователей от этих списков могут варьироваться в зависимости от того, где представлена информация, например, таблицы, словари, телефонные справочники и т.д.



Формат адреса

Здесь стоит быть максимально внимательным и проявлять гибкость при проверке данных. Учитывайте то, что нету универсального порядка записи адреса. В разных странах способ написания может отличаться, к примеру где-то это может быть "Страна/Город/Улица/Дом/Квартира", а где-то "Улица/Дом/Квартира/Страна/Город". Также некоторые данные могут опускаться либо наоборот, добавляться, в зависимости от страны.

Крайне важно следить за способом ввода адреса, особенно это связанным с магазинами, доставкой, логистикой и т.п.

Формат телефонных номеров

Здесь разниться количество чисел, их группировка, использование дополнительных символов и символов-разделителей, таких как дефис (-), точка (.), знак международного префикса (например знак плюс (+)), круглые открывающие и закрывающие скобки (), пробел () и т.д. Обычно для записи номера телефона используется 7-11 цифр.

Числовые и цветовые ассоциации

Такие региональные и культурные особенности как числовые и цветовые ассоциации крайне важно удерживать во внимании при их использовании.

Это очень важный пункт.

Рекламные ограничения

С рекламой всё тоже не так однозначно. Одни и те же вещи могут иметь различные возрастные ограничения в разных странах. Что-то так вообще запрещается рекламировать в одном регионе/стране, но без проблем можно в другом.

Звуки

Звуки, как и музыкальные ассоциации, тоже могут быть интерпретированы по разному. К примеру часто различается текстовое и звуковое использование звуков животных, природы, механизмов и т.д.

Заключение

Как вы видите, тестирование интернационализации и локализации — это интересное и комплексное занятие, на которое должно уделяться приличное время. Это очень важные аспекты современных приложений для запуска на глобальном рынке! Локализация не заключается в простом переводе текста с одного языка на другой. При этом заниматься тестированием

интернационализации и локализации занимаются QA инженеры, многие аспекты можно проверить без (глубокого) знания языка локали! Очень важно иметь четкие требования!

Приведенные мною примеры - это лишь часть! Можно дополнить ещё большим количеством менее распространённых проверок для частных случаев, например лицензии, карты, обработка кредитных карточек и т.д.