

Сравнительный анализ форматов wiki-разметки

К. А. Маслинский
kirill@altlinux.ru

ALT Linux

Третья конференция разработчиков свободных программ
на Протве
Обнинск, 14–26 июля 2006

Обзор

- 1 Зачем сравнивать wiki
- 2 Как это следует делать
- 3 Некоторые результаты

Outline

- 1 Зачем сравнивать wiki
- 2 Как это следует делать
- 3 Некоторые результаты

Попытки стандартизации wiki

Различные вики-форматы весьма похожи но всегда неодинаковы. В последнее время разработчики wiki активно пытаются найти способы сократить «избыточное» многообразие в этой области.

- Попытки сформулировать стандартный язык/подмножество элементов разметки (wiki standards working group);
- Попытки найти формат для обмена данными между разными wiki (Wiki interchange format);
- Попытки создать единый программный инструментарий (библиотеку) для разработки wiki-парсеров.

Так или иначе, все эти попытки упираются в необходимость установить универсальные и специфические черты в форматах wiki-разметки.

Форматы wiki

- UseMod
- DokuWiki
- MediaWiki
- WackoWiki
- TWiki
- MoinMoin
- FlexWiki
- GoWiki

Осложняющие обстоятельства

Сравнение осложняется следующими обстоятельствами:

- отсутствуют описания синтаксиса разных wiki в какой бы то ни было стандартной форме;
- даже существующие нестандартные описания часто неполны и не структурированы;
- парсеры, являющиеся реализацией синтаксиса, также нестандартны и нечитабельны.

Outline

- 1 Зачем сравнивать wiki
- 2 Как это следует делать
- 3 Некоторые результаты

Экскурс в теорию

Рассуждение об универсальном в синтаксисе wiki часто ведётся в терминах обязательных/факультативных **элементов** (тегов).

Язык разметки — это знаковая система, в которой элементы приобретают значение **в противопоставлении друг другу**.

Поэтому сравнивать форматы нужно:

- не поэлементно, а как системы противопоставленных элементов;
- учитывать общие противопоставления, а не только общие элементы.

Методы лингвистической типологии:

- Универсалии
- Импликативные универсалии

Общие элементы

- абзац
- курсив
- полужирный
- нумерованный список
- нумерованный список
- заголовки
- горизонтальная линия
- гиперссылка
- именованная гиперссылка

Outline

- 1 Зачем сравнивать wiki
- 2 Как это следует делать
- 3 Некоторые результаты

Проблемы сопоставления форматов

- Элементы, обусловленные видом тегов
 - `many CamelWord""s`
 - продолжение элемента списка (больше строки)
- Разный уровень детализации
 - экранирующие окружения типа verbatim: пробелы, вики-разметка, html-разметка, html entities, всё
- Незамкнутость формата (HTML)

Характеристики wiki-разметки

- Тенденция к выделению и детализации подсистем
- Баланс и эквивалентность на уровне подсистемц
- Образование «окна» в позиционирование элементов средствами HTML
 - таблицы
 - рисунки
 - текст

Заключение

Всё не так просто...