

GOSTdown – средство автоматической вёрстки и оформления научных отчётов, программной документации и Трудов ИПА РАН

Д. А. Павлов, А. Г. Водолагина



Научный семинар ИПА РАН
12.12.2018

Предпосылки

Заказчики по ОКР и НИР требуют предоставления материалов в формате Word.

Перед научными сотрудниками всех рангов стоят следующие задачи:

- Нумерация глав и разделов, генерация оглавления
- Вставка и нумерация формул, таблиц и рисунков
- Нумерация и форматирование многоуровневых списков
- Ссылки на главы, разделы, формулы, таблицы и рисунки
- Ведение перечня использованных источников, создание списка литературы и ссылок на него в тексте
- Подсчёт и вставка в текст количества страниц, таблиц, рисунков, приложений
- Вставка текста из других документов и оформление его под стиль
- Слияние изменений, сделанных разными людьми в своих копиях документа
- Соблюдение ГОСТ
- Выбор санатория для восстановления после сдачи отчёта

Проблемы редактирования DOCX

- Необходим ПК с Word (Libreoffice/Openoffice, прочие десктопные аналоги и мобильные программы годятся только для просмотра)
- Невозможно откатить или просто просмотреть изменения, сделанные при прошлом редактировании
- Отслеживание изменений затруднено неотделимостью содержания от оформления
- Нет средства централизованного контроля изменений (оно существует и называется Sharepoint, но не установлено)
- Пользователи не могут себя заставить работать строго по стилям; ручная правка форматирования приводит к невозможности централизованного изменения оформления
- Пользователи никогда не научатся средствам Word для автонумерации формул, таблиц и рисунков и создания ссылок на них в тексте; ручная нумерация приводит к потере целостности при удалении/добавлении текста.

Решение

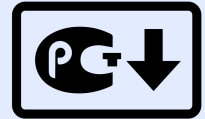
1. Пользователям не нужно работать с оформлением
2. Работа с содержательной частью документа ведётся в языке разметки в любом текстовом редакторе
3. Создание файла в формате DOCX делается автоматической процедурой с соблюдением требований к оформлению
4. Отслеживание и слияние изменений делается через Git

Кроме текста:

- Таблицы (делаются средствами разметки)
- Формулы (аналогично)
- Картинки (отдельные файлы PNG, JPEG, EPS/EMF)
- Библиография (отдельный файл BibTeX)

GOSTdown

GOSTdown = ГОСТ + Markdown



ГОСТ 7.32 – стандарт оформления отчётов о научно-исследовательской работе

ГОСТ 19 – Единая система программной документации (ЕСКД)

ГОСТ 2 – Единая система конструкторской документации (ЕСПД)

История языков разметки:

- troff (1971)
- **TeX (1978)**
- Texinfo (1986)
- **HTML (1993)**
- Wiki (1994)
- AsciiDoc (2002)
- **Markdown (2004)**

Базовые средства Markdown:

Заголовок

Подзаголовок

****жирный**** *курсив* S~22~ c^2^

* маркированный

* список

1. нумерованный

1. список

Почему Markdown?

Pandoc – универсальный преобразователь форматов документов

Поддерживает десятки форматов, включая LaTeX и DOCX.

«Pandoc Markdown» – диалект Markdown с дополнительными возможностями:

- Формулы в формате LaTeX
- Таблицы, картинки в форматах PNG/JPEG/EPS/EMF
- Моноширинный исходный код с раскраской, вложенные списки
- Автонумерация разделов, формул, таблиц, рисунков, ссылки по идентификаторам
- Библиография в формате BibTeX

Все другие программы и форматы далеко отстают по возможностям.

Почему BibTeX?

Нет нужды оформлять список литературы вручную

BibTeX – простой текстовый формат библиографической записи, стандартный для TeX, LaTeX, Pandoc и ряда других программ.

Нет нужды создавать записи в BibTeX-файле вручную

Записи с формате BibTeX предоставляют сайты:

- Журналов и издательств
- ИПА РАН (для Трудов ИПА РАН)
- NASA ADS
- ResearchGate
- <https://www.doi2bib.org/>

Процедура компиляции DOCX

1. Вызов Pandoc для преобразования Markdown в DOCX
 - Используется вспомогательный DOCX-файл со стилями и рамочками (есть образцы)
 - Картинки из внешних файлов внедряются в документ
 - Используется файл формата BibTeX с библиографией и файл формата CSL (готовый) для задания стиля оформления списка литературы
2. Постобработка DOCX с помощью скрипта на Powershell (готового)
 - Придывается титульный лист и лист регистрации изменений из вышеуказанного вспомогательного DOCX-файла
 - Дорабатываются стили для соответствия ГОСТ или стилю Трудов ИПА
 - Внедряются шрифты
 - Сохраняется DOCX и PDF

При использовании Git компиляция может вестись на удалённом сервере.

Использование Markdown и Git в ИПА РАН

- Статьи (Markdown/LaTeX, Git)
- Программный код (Git)
- Программная и конструкторская документация по ОКР (Markdown, Git)
- Отчёты по НИР, годовые отчёты (Markdown, Git)
- Содержимое сайта (Markdown, Git)

Благодаря Git:

- Резервное копирование
- Полная история изменений
- Параллельное редактирование и автоматическое слияние изменений
- Всегда ясно, где лежит свежая версия документа/программы
- Возможность редактирования в онлайн

Благодаря единому формату:

- Лёгкость переноса текста между статьями, отчётами и сайтом

Как использовать GOSTdown в ИПА РАН

1. Создайте репозиторий в Git
2. Скопируйте файлы из репозитория с НТО, программной или конструкторской документацией или статьёй в Труды
3. Удалите всё лишнее
4. Отредактируйте титульную страницу
5. Редактируйте Markdown-файл в Far, Notepad++, любом другом редакторе
6. Отдавайте изменения в Git
7. Скачивайте DOCX и PDF

или

1. Попросите одного из коллег разобраться с пп. 1-4 и 6-7
2. Занимайтесь п. 5

Публично доступные образцы с подробными инструкциями:

<https://gitlab.iaaras.ru/iaaras/gostdown>

<https://gitlab.iaaras.ru/iaaras/transactions-template>

Внутренние образцы: лежат в Git, спрашивайте коллег

Пример с предпросмотром (VS Code)

```
## Таблица с ячейками из многих строк
-----
|   | Номер | Описание параметра |
|---|---|---|
| 188-190 | Углы  $\varepsilon_x$ ,  $\varepsilon_y$  и  $\varepsilon_z$  ориентации эфемерид в ICRS |
| 201-700 | Гравитационные параметры ( $Gm$ ) планет и астероидов |
| 701-706 | Элементы орбиты Меркурия:  $\ln(a)$ ,  $\sin i \cdot \cos(\Omega)$ ,  $\sin i \cdot \sin(\Omega)$ ,  $\cos(\varpi)$ ,  $e \cdot \cos(\varpi)$ ,  $e \cdot \sin(\varpi)$ ,  $l$ , где  $a$  — большая полуось,  $i$  — наклон орбиты,  $\Omega$  — долгота восходящего узла,  $e$  — эксцентриситет,  $\varpi$  — долгота перигелия,  $l$  — средняя долгота |
| 707 |  $\dot{a}/a$  Меркурия |
-----

Table: Таблица с многострочными (но одноабзацевыми) ячейками.
{#tbl:mytable-multiline}

## Таблица с ячейками из многих абзацев
: Таблица с многоабзацевыми ячейками. {#tbl:mytablegrid}

+-----+-----+-----+
| Колонка 1 (R) | Колонка 2 (C) | Колонка 3 (L) |
+-----+-----+-----+
| Просто абзац | Просто текст | - Маркированный |
| текста на | | - список с длинным |
| помещающимся на одну строку | | текстом, не |
| несколько строк. | | |
| И за ним ещё один. | | i. Вложенный |
| | | i. Нумерованный |
| | | i. Список |
| | | - Пункт списка с формулой внутри:  $\int_{a}^b x^2 dx$ . |
+-----+-----+-----+
| Отдельная формула: | | 1. `НУМЕРОВАННЫЙ` | |
|  $e^{\pi i} + 1 = 0$  | | 2.  $s + \pi + sok$  |
| | |  $::: \{custom-style="Normal"\}$  |
| 1. Разреженный | | Новый абзац текста после нумерованного |
| списка. | | |
| | | | Ему принудительно выставлен стиль «Normal» |
| (Обычный); из-за внутренних особенностей Pandoc | | |
| 2. Нумерованный | | | всем абзацам ячейки таблицы, где есть |
| компактные списки, присваивается по умолчанию стиль «Compact». | | |
| | | | Различие в данном случае только в
```

Границы допускаемой абсолютной погрешности измерений ЯТ по каналам А и Б, К, не более	±2,5
---	------

Таблица с ячейками из многих строк

Table 5: Таблица с многострочными (но одноабзацевыми) ячейками.

Номер	Описание параметра
188-190	Углы ε_x , ε_y и ε_z ориентации эфемерид в ICRS
201-700	Гравитационные параметры (Gm) планет и астероидов
701-706	Элементы орбиты Меркурия: $\ln(a)$, $\sin i \cdot \cos(\Omega)$, $\sin i \cdot \sin(\Omega)$, $e \cdot \cos(\varpi)$, $e \cdot \sin(\varpi)$, l , где a — большая полуось, i — наклон орбиты, Ω — долгота восходящего узла, e — эксцентриситет, ϖ — долгота перигелия, l — средняя долгота
707	\dot{a}/a Меркурия

Таблица с ячейками из многих абзацев

Table 6: Таблица с многоабзацевыми ячейками.

Колонка 1 (R)	Колонка 2 (C)	Колонка 3 (L)
Просто абзац текста на несколько строк. И за ним ещё один.	Просто текст	<ul style="list-style-type: none"> Маркированный список с длинным текстом, не помещающимся на одну строку <ol style="list-style-type: none"> Вложенный Нумерованный Список Пункт списка с формулой внутри: $\int_a^b x^2 dx$.
		<ol style="list-style-type: none"> НУМЕРОВАННЫЙ $s + \pi + sok$

Ограничения и дополнительные возможности

Ограничения

- Нет поддержки таблиц с объединёнными ячейками (но когда-нибудь появится)
- Невозможно сделать разное выравнивание в заголовке таблицы и в ячейках под ним
- В режиме предпросмотра в VS Code не видны векторные картинки (EPS/EMF)
- В итоговом DOCX номера в ссылках являются простым текстом (но с гиперссылкой)
- Нет поддержки ссылок между документами

Дополнительные возможности (требуют усилий от пользователя)

- Составление документа из нескольких Markdown-файлов
- Вставка произвольных DOCX-файлов в произвольные места документа
- Указание специальных стилей форматирования для фрагментов текста

Демонстрация