

Инструкция-протокол разметки телеграмм

Инструкция разбита на категории; каждая из категорий представляет собой колонку в базе данных с разметкой телеграмм.

У нас существует некоторая иерархия столбцов в базе данных; каждый из них выделен особым цветом.

- **Столбцы зелёного цвета:** это данные, которые добыты автоматически, в них **не надо вносить изменения**; в некоторых случаях из них можно черпать информацию для разметки остальных колонок.
- **Столбцы красного цвета:** это так называемые “сырые” столбцы, т.е. те данные, поля для которых есть в теле телеграммы. Их можно и нужно размечать.
- **Столбцы синего цвета:** это данные, которые добыты искусственным путём, поскольку их нет в теле телеграммы. Например, в телеграмме нет специального поля “тема” – это что-то, что мы можем вывести на основе содержания телеграммы.

1. ID

ID — уникальный ключ в нашей локальной базе данных. Он состоит из пометки “tg”, *номера телеграммы и номера стороны*. Номера телеграмм начинаются с 00001.

Номера сторон: если телеграмма односторонняя, то номер автоматически записывается как 1. Если двусторонняя, то 1 — сторона с текстом, 2 — сторона с изображением.

Пример: tg00001-1, если это скан телеграммы со стороны с текстом;
tg00001-2, если это скан телеграммы со стороны с изображением.

NB! В таблицу записывается только номер: tg00001.

2. ID Госкаталог

Это уникальный ключ объекта из базы данных Госкаталога, откуда был взят текущий корпус телеграмм. Он нужен для привязки нашей базы данных к базе данных Госкаталога, поскольку в нашей БД мы делаем какие-то изменения (например, удаляем неподходящие нашим целям телеграммы).

3. Ссылка

- **Если это корпус Госкаталога:**

Это ссылка из базы данных Госкаталога. Если вместо ссылки стоит “NA”, значит, для телеграммы нет изображения, и, соответственно, текста телеграммы тоже.

- **Если это собственный корпус:**

Здесь должна быть ссылка на файл на Гугл-диске в папке [Телеграммы \(сканы\)](#). У файла должно быть название, совпадающее с ID. Если телеграмма взята из какого-либо источника в интернете, необходимо вставить ссылку на первоисточник.

4. Текст

Для этой графы необходимо максимально точно перепечатать то, что написано в теле телеграммы, сохраняя характерную стилистику телеграмм (например, пропуск предлогов, знак равно (“=”) перед подписью отправителя и т. д.). Текст начинается после служебного заголовка.

Также необходимо сохранять регистр написания телеграмм. Если телеграмма печатная, и в ней, как правило, текст написан капслоком, то при разметке текст тоже должен быть внесён в таблицу капслоком. Если телеграмма рукописная, регистр телеграммы при разметке должен быть таким же, как в оригинальной телеграмме.

Если же что-то написано неразборчиво, это необходимо обозначить как *(нрзб)*.

NB! Есть несколько нюансов:

1. Если телеграмма написана согласно нормам **дореволюционной орфографии** (до реформы 1918 года), то необходимо сначала отразить оригинальное написание текста, а затем в фигурных скобках {...} написать это же слово, но уже на современный манер. То есть, это должно выглядеть примерно так:

подлежит удержанию в средства профессионального {профессионального} союза...

Буквы, характерные для дореволюционной азбуки, также необходимо отразить их оригинальном написании: І і, Ъ ъ, Ѳ ѳ, V v.

2. Если в тексте есть очевидные **опечатки**, то нужно сначала отразить исходный вариант написания, затем в угловых скобках <...> отразить правильный вариант написания слова, к примеру:

ПОЗДРАВЛЯЮ РАДОСТНЫМ СЧСТЛИВЫМ <СЧАСТЛИВЫМ> ДНЕМ РОЖДЕНИЯ

3. Если в тексте телеграммы есть **орфографические ошибки**, нужно сначала отобразить оригинальный вариант написания, затем в звёздочках *...* написать орфографически верный вариант написания того или иного слова, например:

*АНТИПАРТИЙНОЙ ГРУППОЙ ЗИНОВЬЕВСКОЙ АППОЗИЦИИ *ОППОЗИЦИИ**

Однако следует дифференцировать между орфографическими ошибками и дореволюционными орфографическими нормами. К примеру, слово типа

“профессионального” будет считаться дореволюционной орфографией, подробнее см. в [Википедии](#).

4. Если в тексте телеграммы есть аббревиатуры и сокращения, то нужно сперва написать оригинальный вариант написания слова, затем под такими ^...^ знаками написать расшифровку. НО это не касается общепринятых аббревиатур типа СССР, ООН и т. д.

ЗДРАВСТ /ЗДРАВСТВУЙ/ МИЛАЯ

5. Если язык телеграммы не совпадает с графикой (например, телеграмма на русском языке, но транслитом, или на украинском языке, но без использования украинских букв, следует пометить это так: “оригинальный текст & текст с нормальной графикой&”
6. Если текст телеграммы написан **неразборчиво**, не нужно пытаться всеми силами разобрать эти кракозябры, просто напишите (**нрзб**) – “неразборчиво”.
 - Если неразборчива часть слова, следует прописать то, что понятно, а остальное пометить как (нрзб), например:
Викторова при(нрзб) необходим.
 - Если неразборчиво несколько слов подряд, напишите вместо них одно (*нрзб*). То же самое касается тех случаев, когда неразборчив весь текст телеграммы.
7. Если в телеграмме имеются **фактические ошибки**, известные только разметчику, то необходимо сперва отразить оригинальный вариант написания, после этого в квадратных скобках [...] написать фактически верный вариант. Например:

СТАНИСЛАВУ [ВЯЧЕСЛАВУ] ПЕТРОВИЧУ [ПАНТЕЛЕЕВИЧУ]

8. Если текст телеграммы **отсутствует** – следует написать **NA** в поле текста.

Данные правила касаются не только колонки “Текст”, но и всех остальных нижеупомянутых колонок, поскольку опечатки, дореволюционная орфография, неразборчивый текст и т. д. могут встречаться везде.

Также, несмотря на то, что в конце текста телеграммы после знака равно (“=”) чаще всего идёт подпись отправителя, и для этого существует отдельная колонка, подпись отправителя после знака равно также нужно отразить в колонке “Текст”.

5. Название

В этой колонке представлены названия телеграмм согласно базе данных Госкаталога. Здесь категорически нельзя ничего менять, поскольку это очень важная колонка для сбора информации о телеграмме. Например, если телеграмма рукописная, и почерк очень непонятный, из этой колонки можно понять, о чём речь, или хотя бы кто кому писал.

6. Отправитель

В большинстве случаев в телеграммах не предусмотрено отдельное поле для данных отправителя, и отправитель чаще всего подписывается неформально в конце текста телеграммы после знака равно (“=”), например “мама”, “папа”, “Петя” и др. Именно поэтому данная колонка выделена синим цветом – для неё мы добываем информацию косвенным путем.

- Если есть имя отправителя: пишем в формате Фамилия Имя Отчество
- Если отправитель не указан: пишем *NA*
- Если есть только неформальная подпись: пишем *NA* (неформальная подпись), например: *NA (Наташа)*
- Если телеграмма от коллектива людей: пишем *коллективный (семья Дроздовых)*, или *коллективный (коллектив Погрузнинской МТС)*
- Если телеграмма от организации: пишем *организация (Погрузнинская МТС)*

NB! Часто бывает так, что телеграмма одновременно и от определенного лица, и от организации/коллектива. В таком случае нужно просто указать через запятую отправителей, например: *Дроздов, организация (ЦК КПСС)*.

7. Получатель

Эта колонка предназначена для данных, связанных с получателем телеграммы. Данные получателя телеграммы указываются сразу после места получения.

- Если есть имя получателя: пишем в формате Фамилия Имя Отчество
- Если получатель не указан: пишем *NA*
- Если телеграмма предназначена коллективу людей: пишем *коллективный (семья Дроздовых)*, или *коллективный (коллектив Погрузнинской МТС)*
- Если телеграмма предназначена организации: пишем *организация (Погрузнинская МТС)*

8. Место отправления

Колонка для указания места отправления телеграммы. Место отправления указывается в начале телеграммы до служебного заголовка. Мы фиксируем *исторические географические данные*, т.е. территориальную принадлежность места на момент отправления телеграммы. Так, если телеграмма была отправлена из Москвы в 1955 году, место отправления будет таким: *РСФСР, г. Москва*.

Формат указания географических данных: *страна, область, район, населенный пункт*. Если населенный пункт административного значения (т.е. название города совпадает с названием области или района), то писать область и район не обязательно.

Если место отправления неизвестно, пишем *NA*.

9. Место получения

Место получения телеграммы указывается после служебного заголовка и серии телеграммы. Мы фиксируем *исторические географические данные*, т.е. территориальную принадлежность места на момент отправления телеграммы. Так, если телеграмма была получена в Куйбышеве в 1986 году, место отправления будет таким: *РСФСР, г. Куйбышев*, несмотря на то, что сейчас этот город называется Самара.

Формат указания географических данных: *страна, область, район, населенный пункт*. Если населенный пункт административного значения (т.е. название города совпадает с названием области или района), то писать область и район не обязательно.

Если место получения неизвестно, пишем *NA*.

10. Описание

Это дополнительная колонка из Госкаталога. В ней гораздо меньше потенциально полезной информации, чем в столбце “Название”, но иногда из нее все же есть, что почерпнуть. Впоследствии она будет объединена с колонкой “Название” и перенесена в поле “Примечание”.

11. Дата

Точная датировка телеграмм на данный момент представляет собой большую проблему. В подавляющем большинстве телеграмм указывается только *день, месяц и время* отправки телеграммы, но не год.

Если год отправки телеграммы еще не указан, его можно узнать либо из колонки “Название”, либо по косвенным признакам:

- Год отправки телеграммы может быть указан на штемпеле (печати), но его далеко не всегда можно разобрать, поскольку печать часто смазывается.
- На бланке телеграммы/открытки может быть указан тираж и год выпуска. Указанный год — минимальная дата отправки телеграммы, т. е. телеграмма была отправлена не раньше указанного года. Это крайне ненадежная метрика, поскольку временной промежуток между выпуском бланка телеграммы/открытки и отправкой самой телеграммы может составлять десятки лет, но за неимением лучшего варианта можно указать хотя бы таким образом.

Дата телеграммы указывается в формате ДД.ММ.ГГГГ. Если что-то из этого неизвестно, то цифра заменяется латинской буквой *x* (например, если неизвестен год — *31.05.xxxx*; если известен приблизительный год, можно указать его в таком формате — *31.12.200x*).

12. Дата_расчищено

Эта колонка предназначена для наших попыток унифицировать уже имеющиеся даты телеграмм автоматически. Впоследствии предыдущая колонка будет заменена этой.

13. Время приёма

Время приема телеграммы указывается в служебном заголовке сразу после даты. Формат разметки времени: ЧЧ:ММ.

МОСКВА 112641/13 23 15/4 1640=, где 16 — часы, 40 — минуты.

Если неизвестно время отправления телеграммы, нужно писать *NA*.

14. Время передачи

Время передачи телеграммы указывается либо в левом верхнем углу телеграммы, либо в соответствующем поле телеграммы “Прием”. Формат разметки времени: ЧЧ:ММ.

Если неизвестно время отправления телеграммы, нужно писать *NA*.

15. Служебное

Эта колонка предназначена для служебного заголовка, который находится в телеграмме после места отправления. Служебный заголовок содержит данные, необходимые для опознавания телеграммы. Он выглядит следующим образом:

МОСКВА 112641/13 23 15/4 1640=

- *112641/13* — исходящий номер телеграммы;
 - ❖ *112641* — номер абонентской установки;
 - ❖ *13* — порядковый номер телеграммы с начала суток;
- *23* — количество слов;
- *15/4* — день и месяц отправки телеграммы;
- *1640 =* — время отправки телеграммы.

Несмотря на то, что для даты и времени отправления телеграммы предусмотрены отдельные колонки, их также необходимо прописать в служебном заголовке.

Если телеграмма на бланке открытки, также в этой колонке нужно указать серию телеграммы. Она может быть расположена чуть ниже служебного заголовка, на уровне начала текста телеграммы. Серия состоит из буквы и двузначного номера через дефис или без него, например, *E-52*.

Написано неразборчиво: **(нрзб)**

Не указано: **NA**

Дореволюционная орфография: {...}

Опечатки: <...>

Орфографические ошибки: *...*

Сокращения и аббревиатуры: ^...^

Несовпадение языка и графики: &...&

Фактические ошибки: [...]

16. Язык

В этой колонке указывается язык, на котором написана телеграмма. Уже имеющиеся языки в нашем корпусе: русский, украинский, английский, немецкий, болгарский.

Если язык неизвестен (т.е. поле “Текст” отсутствует), пишем *NA*.

17. Графика

В этой колонке указывается графика написания телеграммы. Графика может быть либо латиница, либо кириллица. Иногда бывает так, что телеграмма написана на русском языке, однако графика латинская – именно для таких случаев введена эта колонка.

NB! В таких случаях в поле “Текст” нужно сначала отразить оригинальный вариант написания телеграммы, а потом под вот такими вот &...& символами перепечатать текст на кириллической графике, например:

Privet vsem eto ya telegramma &привет всем это я телеграмма&

Если графика неизвестна (т.е. поле “Текст” отсутствует), пишем *NA*.

18. Вид бланка

Бланк телеграммы может быть представлен в двух вариантах: *рукописный* и *печатный*. Иногда на телеграмме может быть одновременно и рукопись, и печатный текст; в таком случае необходимо через запятую написать эти варианты: *печатный, рукописный*.

19. Вид телеграммы

Телеграмма может быть следующих видов:

- Обычная;
- Открытка
- Срочная;
- Местная;
- Международная;
- Почто-телеграмма;
- Копия;
- Радиотелеграмма;

- Фототелеграмма;
- Правительственная;
- Военная;
- Транзитная;
- Авиа и др.

Обычно это написано в самом верху телеграммы на бланке, на печати или в служебном заголовке.

20. Тема

В данной графе указывается тема телеграммы; то, о чём говорится в сообщении. На данный момент у нас уже сформировался определённый пул тем, однако можно добавлять свои:

1	присвоение звания	177
2	получение награды	96
3	день рождения	44
4	новости	38
5	просьба	22
6	поздравление	18
7	уведомление	13
8	новости, просьба	10
9	требование	8
10	соболезнование	8
11	новый год	7
12	годовщина октябрьской революции	6
13	пожелание	5
14	приглашение	4
15	день города	4
16	открытие съезда комсомола	4
17	благодарность	4
18	первое мая	3
19	поздравление, получение награды	3
20	перевыполнение плана	3
21	новости, благодарность	2
22	новости, пожелание	2
23	девятое мая	2
24	благодарность, обещание	1
25	день рождения, годовщина октябрьской революции	1
26	приветствие	1
27	восемь марта	1
28	новости, поздравление	1
29	день города, получение награды	1
30	присвоение звания, приглашение	1
31	рождение ребенка	1
32	получение награды, приглашение	1
33	уведомление, просьба	1
34	одобрение	1
35	день советской армии	1

21. Тон

В этой графе нужно указать, какой у сообщения тон. Он бывает *позитивный*, *негативный*, *нейтральный* или *смешанный*.

- *Позитивный* — это хорошие новости: рождение ребёнка, поздравления
- *Негативный* — плохие новости, известие о смерти, сочувствие
- *Нейтральный* — когда эмоциональная окраска отсутствует
- *Смешанный* — писать в тех случаях, когда однозначно тональность определить невозможно.

22. Персоналии

Данная колонка предназначена для упоминаний известных личностей — деятелей политики, культуры, спорта и т. д. Персоналии указываются в формате *Фамилия Имя Отчество*.

Если персоналия — это отправитель или получатель телеграммы, пишем только её ФИО. Если известная личность упоминается в тексте телеграммы, пишем *ФИО (упоминание)*.

23. Источник

В этой колонке необходимо указать источник телеграммы. Можно выбрать одну из предложенных опций или добавить свою:

- *Госкаталог* — если телеграмма из базы данных госкаталога;
- *Семейный архив* — если телеграмма взята из семейного архива;
- *Музей* — если телеграмму предоставил музей;
- *Форум коллекционеров* — если телеграмма взята из форума коллекционеров соответственно.

24. Примечание

В данной колонке указывается любая информация, отсутствующая в источнике и известная только разметчику.

Если у вас остались какие-то вопросы, пишите нам в Telegram: @purrmorning, @arstotrix или @therealcreogenic.