**Рекомендации для составителей вопросов:**

Главная рекомендация для составителей вопросов заключается в том, что вопросы должны как можно меньше пересекаться. Уточняющие вопросы должны задаваться в блоке с основным вопросом. Все блоки вопросов должны быть пронумерованы. Это сделано специально, чтобы программе не приходилось по тексту вопроса определять к какому вопросу он относится, опыт практики показал, что при соотнесении текста с вопросом из таблицы возникает недопустимо много ошибок. У этого есть две основные причины:

1. Несовершенство нейросетей. В любом использовании нейросетей должно закладываться права на ошибку.
2. Человеческий фактор. Первая причина сильно усугубляется тем, что интервьюеры любят перефразировать вопрос настолько, что алгоритмы типа fuzzy search уже не могут определить к какому вопросу из плана интервью относится вопрос, который задает сейчас интервьюер

Для улучшения качества кодирования, вопросы должны быть самодостаточными, чем меньше связи между вопросами, тем лучше. Это позволит минимизировать ситуации в которых респондент, отвечая на один вопрос, ответить еще на несколько других. Такие ситуации будут сложно разрешимы программой и могут привести к непредсказуемым результатам.  
  
**Рекомендации для интервьюеров:**

При разговоре с респондентом старайтесь не пропускать вопросы из списка. Программа рассчитывает, что вопросы будут задаваться последовательно. В случае, когда вопрос пришлось пропустить, обозначьте к какому блоку вопросов вы приступили. Это упростит работу транскриптору.

Если интервью подразумевает определенное ветвление в зависимости от ответов респондента. Явно проговаривайте, что вы переходите к блоку вопросов под номером N. Это может показаться странным, но такой прием сильно облегчит жизнь транскриптору.

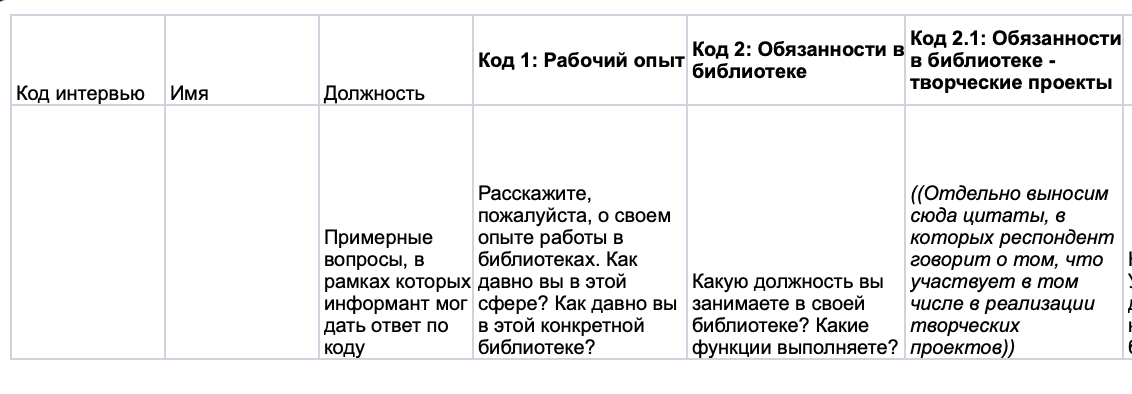
Программа написана таким образом, что ей не важно количество интервьюеров, но желательно, чтобы респондент был один. Если респондентов больше одного, то в самом начале файла с транскриптом, с помощью специальной разметки должно быть описано как программа должна учитывать ответы респондентов (как одна строка в таблице или под одной строке для каждого респондента)  
  
**Рекомендации для транскрипторов:**

Транскриптору разрешается использовать инструменты, которые помогают перевести интервью из аудио/видео формата в текстовый. Однако полученный текст придется разметить. Разметка текста — главная задача транскриптора, если он плохо справится с этой задачей, программа выдаст некорректные результат или вовсе не будет работать. В таком случае, текст придется анализировать вручную.  
  
Важно понимать, что такая ситуация — это не всегда следствие ошибок транскриптора. Чем меньше ошибок допустят составитель вопросов и интервьюер, тем легче будет работать транскриптеру. При этом на транскрипторе лежит ответственность равная, если не превышающая совместную работу составителя и интервьюера.

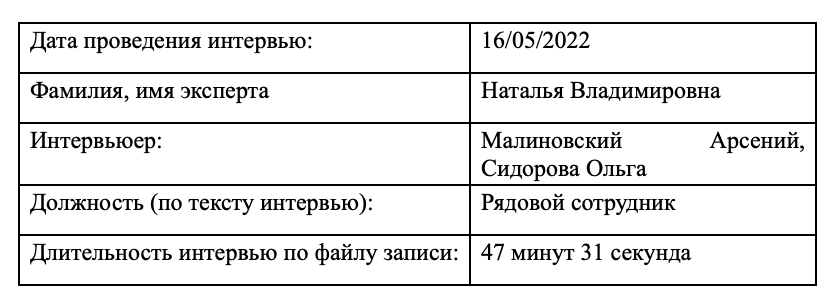
Главная задача транскриптора быть внимательным и педантичным при разметке текста.

**Структура разметки:**

Шапка итоговой таблицы может выглядеть так. Шаблон таблицы тоже загружается человеком в программу. После чего программа проанализирует все docx файлы в папке interviews и заполнит таблицу.



Из шапки видно что в таблице есть обязательное поле **Код интервью**, которое берется из **название docx файла без расширения** (на данный момент программа поддерживает только этот формат)

Дальше идет ряд необязательных полей. В данном примере это Имя и Должность. Количество опциональных полей определяется по началу файла с интервью. Например, если файл расшифровки начинается с такой таблицы, то в шаблоне, после Кода интервью будет еще 5 опциональных колонок и только потом начнутся блоки с вопросами.  
  
Существует только один сценарий, когда таблица с опциональными колонками находится не в начале. Когда респондентов больше чем один. При этом число интервьюеров программу не интересует. Если респондентов больше чем 1. То в начале файла нужно сделать маппинг, который поможет программе соотнести где какой респондент и сделать под каждого респондента отдельную строку в таблице. Наименование Кода интервью при этом немного изменится, чтобы в таблице не было двух одинаковых кодов интервью. N\_<код интервью>, где N — это номер респондента.

Маппинг должен начинаться и заканчиваться символами !#

Например:

!#

Р1 Натали Портман

Р2 Кира Найтли

!#

таким образом в строке в таблице где код интервью 1\_<код интервью> будут ответы Натали Портман, а в строке 2\_<код интервью> будут ответы Киры Найтли

Если же маппинг не сделать, программа будет ожидать, что респондент один и строка в таблице тоже одна. В самом тексте интервью ответы респондента нужно помечать буквой Р (без цифры), а вопросы интервьюера или интервьюеров всегда буквой И (программа не учитывает, что их может быть много).

Дальше идут колонки с блоками вопросов.

Очень важно проверить, что суммарное количество колонок совпадает с количествои необязательных полей + количеством блоков с вопросами + 1 (для обязательной колонки с кодом интервью). Если программа обнаружит несоответствие количества колонок между шаблонной таблицей и файлом интервью, то приоритет отдается шаблону и файл с интервью просто пропускается, в лог будет написана причина, по которой программа пропустила файл.

При повторном запуске программы, будут проанализированы только те интервью, которые до этого еще не анализировались или при анализе которых в прошлый раз возникла ошибка.

После таблицы с опциональными полями в начале файла идут блоки с вопросами.

Начало блока помечается с новой строки ==N , где N Это номер блока (номер колонки в шаблонной таблице куда будут заносится ответы на вопрос)

после номера блока будет вопрос и ответ респондента. Их может быть больше одного, но сильно много вопросов и ответов в одном блоке делать не рекомендуется, потому что все ответы одного блока склеиваются и анализируются совместно. Кроме главного вопроса в блоке могут быть ТОЛЬКО уточняющие к главному вопросы.

Пример:

==1

И: Как вы относитесь к написанию диплома за одну неделю?

Р: Мне кажется это безумие. Никто не может качественно выполнить работу в такие сроки.

И: Сколько времени нужно чтобы написать диплом?

Р: Понадобится минимум месяц.

==2

И: …

Р: …

Еще раз обращаю внимание на то, что ответы на вопросы одного блока будут склеиваться. В итоге в таблице в колонке для первого блока вопросов будет записано следующее: *Мне кажется это безумие. Никто не может качественно выполнить работу в такие сроки. Понадобится минимум месяц.*

**Дополнительные возможности:**

Список дополнительных возможностей будет дополняться со временем и с целью обратной совместимости будет фиксироваться маркерами. Маркеры выставляются перед новый блоком вопроса. Для каждого блока можно указать только один маркер или не указать вовсе.  
  
На данный момент доступно два маркера:  
  
SA — Sentiment Analysis определяет отношение респондента и возвращает численное представление этого отношения. Например, -1 — резко негативное, 0 — нейтральное, 1 — очень позитивное

KW — Key Words вычленяет из текста ответа ключевые слова

Если маркер не ставить, то все ответы данного блока вопросов будут склеены и отправлены целиком в таблицу.

Маркеры выставляются по согласованию. Наиболее логичным кажется выставлять маркеры на усмотрение транскриптора или, если это заранее спланировано, на усмотрение составителя вопросов. В общем случае, чем больше вопросов помечены маркерами, тем больше информации можно достать из интервью. Оставлять блок без маркера можно только в том случае, когда непонятно какой маркер даст хорошие результаты или когда есть четкая необходимость в полном ответе на вопрос и дальнейшем ручном анализе ответа на этот вопрос.

Пример:

Вот так к ответам на вопросы из первого блока будет применяться анализ тональности текста

==1 SA

И: Как вы относитесь к написанию диплома за одну неделю?

Р: Мне кажется это безумие. Никто не может качественно выполнить работу в такие сроки.

И: Сколько времени нужно чтобы написать диплом?

Р: Понадобится минимум месяц.

==2

И: …

Р: …

А вот так будет применяться высленение ключевых слов  
  
==1 KW

И: Как вы относитесь к написанию диплома за одну неделю?

Р: Мне кажется это безумие. Никто не может качественно выполнить работу в такие сроки.

И: Сколько времени нужно чтобы написать диплом?

Р: Понадобится минимум месяц.

==2

И: …

Р: …