

Гайд для участников: сбор данных для эксперимента

Что мы делаем и зачем

Я провожу эксперимент по сравнению разных способов организации поиска по базам судебной практики. Мне нужны реальные подборки дел и реальные вопросы, которые юристы задают к таким подборкам.

Ваш вклад: подборка судебных актов по теме, в которой вы разбираетесь, + вопросы к ней + разметка (какие дела отвечают на какой вопрос).

Что от вас нужно

1. Подборка судебных актов (15–30 решений)

Откуда брать: - Ваши рабочие подборки по реальным проектам - Подборки из К+ или Гарант, к которым вы обращаетесь и считаете неплохо собранными

Какие дела подойдут: - По любой отрасли права (гражданское, административное, уголовное, налоговое — что угодно) - Неплохо, если в подборке есть дела с **разными исходами** по похожим вопросам или разные варианты фактических обстоятельств, правовой вопрос при которых был разрешен похожим образом - В подборке могут быть решения разных инстанции и по разным регионам

Формат файлов: - Идеально: .txt (просто скопировать текст решения в текстовый файл) - Допустимо: .pdf, .doc, .docx - **Не подходит:** сканы, скриншоты, изображения

Как назвать файлы:

001.txt
002.txt
003.txt
...

Просто ваша фамилия и порядковые номера. Так мне проще всего будет их программно обработать и свести в один датасет. Официальный номер дела (типа А40-12345/2024) можно указать в маппинге в таблице (см. ниже).

Как упаковать: - Все файлы — в одну папку - Папку заархивировать в .zip - Назвать архив своей фамилией на латинице, например: ivanov.zip

2. Вопросы к подборке (5–10 вопросов)

Сформулируйте вопросы, ответы на которые можно найти в вашей подборке.

Хороший вопрос для целей эксперимента должен касаться правовой аргументации суда применительно к какому-то набору фактических обстоятельств. Вопросы могут быть разных типов, особенно приветствуются: - ответ на вопрос есть в одном документе из подборки - для ответа на вопрос нужен синтез позиций из нескольких документов из подборки - вопрос касается какой-то тенденции разрешения типа споров судами / преобладающей позиции, которую можно вычлениить из подборки

Вопросы, ответы на которые и так можно найти в законе, для эксперимента не совсем подойдут.

 **Примеры хороших вопросов:**

- Признаётся ли покупатель вторичной недвижимости добросовестным, если он не пригласил на сделку психиатра?
- При каких обстоятельствах номинальный директор может быть привлечен к субсидиарной ответственности? Что должен доказать директор, чтобы избежать привлечения к ответственности?
- Взыскивается ли административный штраф за нарушение маркировки, если товар не был введён в оборот?

 **Примеры плохих вопросов:**

- «Что такое субсидиарная ответственность?» — это определение, ответ не в судебной практике
- «Как взыскать долг?» — слишком общий вопрос
- «Найти дело А40-12345/2024» — это поиск по номеру, а не по смыслу

Важно: вопрос должен быть таким, чтобы на него могли ответить **несколько** дел из подборки (или хотя бы одно, но не все сразу). И в подборке точно должен содержать ответ на вопрос - даже если не полный, то, как минимум, **релевантный** к формированию собственной позиции.

3. Разметка релевантности

Это самая важная часть. Нужно указать, какие дела из подборки отвечают на каждый вопрос (или релевантны к ответу на вопрос).

Что значит «релевантное дело»: - В нём есть информация, которая помогает ответить на вопрос или сформулировать собственную позицию по кейсу - Это может быть прямой ответ или полезный контекст

Пример разметки:

<i>Вопрос</i>	<i>Релевантные дела</i>
Признаётся ли покупатель вторичной недвижимости добросовестным, если он не пригласил на сделку психиатра?	003, 007, 012
При каких обстоятельствах номинальный директор может быть привлечен к субсидиарной ответственности? Что должен доказать директор, чтобы избежать привлечения к ответственности?	001, 005
Взыскивают ли упущенную выгоду без доказательств мер по минимизации потерь?	002, 008, 015, 019

В колонке «Релевантные дела» указываете номера файлов (те самые 001, 002 и т.д.).

Если на вопрос отвечает только одно дело — это нормально. Если ни одно — переформулируйте вопрос.

Итого: что прислать

1. **Архив с файлами** судебных актов (.zip)
2. **Таблицу** (шаблон ниже) в формате Excel или Google Sheets (просто пришлите ссылку). Таблицу в Excel назовите так же, как архив с файлами.

Шаблон таблицы

Таблица состоит в идеале из двух листов. Второй - очень мне поможет, но если вы загружены, то можно ограничиться только первым. Если вы хотите сделать Google_sheets, скопируйте себе шаблон, размещенный по [этой ссылке](#).

Лист 1: Вопросы и разметка

<i>question_id</i>	<i>question_text</i>	<i>relevant_files</i>
Q01	Снижают ли суды неустойку по ст. 333 ГК, если ответчик не заявлял ходатайство?	002, 007, 012
Q02	Признаётся ли покупатель добросовестным без проверки ЕГРН?	001, 005

Расшифровка колонок:

- question_id — номер вопроса (Q01, Q02...)
- question_text — текст вопроса
- relevant_files — номера релевантных файлов через запятую

Лист 2: Маппинг дел (опционально!)

file_id case_number

001	A40-12345/2024
002	A56-67890/2023
003	2-1234/2024

Расшифровка колонок:

- file_id — номер файла из архива (001, 002...)
- case_number — официальный номер дела

Частые вопросы

Сколько времени это займёт? Если у вас уже есть подборка — около 60-80 минут на подготовку вопросов и минимальное оформление. С полным оформлением (оба листа в таблице и архив с тестами в .txt - до 2х часов.

Можно ли взять подборку из К+ или Гарант? Да, если вы её ранее использовали или проверили сейчас и хорошо ориентируетесь в содержании дел.

Что если я не уверен в релевантности? Размечайте как считаете правильным. Субъективность — это нормально. Главное — ваша экспертная оценка.

Можно ли прислать несколько подборок по разным темам? Да! Каждую — отдельным архивом и отдельной таблицей. Название архива можно указать в формате фамилия_отрасль, например ivanov_bankruptcy.zip

Куда присылать

На email yakunenko.ekaterina@gmail.com

Спасибо за помощь! 🙏