

Приложение 1. Структура данных для извлеченной информации

Этот документ описывает структуру JSON-объекта, используемого для фиксации каждого отдельного события или упоминания, извлеченного из дневниковых записей. Каждый объект представляет собой отдельную семантическую единицу, идентифицированную и проанализированную большой языковой моделью.

Структура данных была спроектирована для фиксации не только самого события, но и критически важных метаданных для последующего исторического анализа: источника информации для автора дневника, контекста, степени уверенности модели и точной привязки к исходному тексту. Все сгенерированные объекты проходят строгую валидацию на соответствие этой схеме с помощью библиотеки Pydantic.

Поле (Field)	Тип данных (Data Type)	Описание (Description)
entry_id	Integer	Уникальный идентификатор исходной записи дневника.
event_id	String (optional)	Идентификатор события из онтологии (knowledge_map.json), позволяющий связать упоминание с типологией событий и реакций автора.
event_name	String	Человекочитаемое название события или аспекта из онтологии.
event_subtype_custom	String (optional)	Дополнительное уточнение типа события, сгенерированное моделью для большей детализации.
description	String	Подробное описание события на основе текста дневника с объяснением его связи с революциями 1848–1849 гг.
date_in_text	String (YYYY-MM-DD) / Null	Дата события, если она явно упомянута в тексте (может отличаться от даты записи).
source_date	String (YYYY-MM-DD)	Дата дневниковой записи, к которой относится событие.
location	String	Географическое место события, как оно дословно упомянуто в тексте.
location_normalized	String (optional)	Нормализованное (стандартизированное) название места события на современном русском языке.

brief_context	String	Краткий исторический контекст, сгенерированный моделью и содержащий 1-2 релевантных внешних факта для верификации.
information_source	String	Источник информации, на который ссылается автор дневника (например, «петербургские газеты», «по слухам»).
information_source_type	String (Enum)	Категория источника информации. Возможные значения: «Официальные источники (газеты, манифесты)», «Неофициальные сведения (слухи, разговоры в обществе)», «Личные наблюдения и опыт автора», «Информация от конкретного лица (именованный источник)», «Источник неясен/не указан».
confidence	String (Enum)	Оценка моделью уверенности в релевантности извлеченной информации. Возможные значения: «High», «Medium», «Low».
classification_confidence	String (Enum)	Оценка моделью уверенности в правильности классификации события по онтологии. Возможные значения: «High», «Medium», «Low».
keywords	List of Strings	Список ключевых слов и фраз из текстового фрагмента, релевантных для индексации и поиска.
text_fragment	String	Точная цитата из дневника, на основании которой была извлечена информация.

Пример 1. JSON-объекты, сгенерированный моделью Gemini-2.5-Pro для записи от 7 января 1849 г.

```
[
  {
    "entry_id": 7,
    "event_id": "REV1848_HUN_MIL_AUS",
    "event_name": "Военные действия (Австрия против Венгрии)",
    "event_subtype_custom": "Успехи австрийских войск",
    "description": "Автор, прочитав петербургские газеты, отмечает успехи австрийских войск в ходе подавления восстания в Венгрии.",
    "date_in_text": null,
    "source_date": "1849-01-07",
    "location": "Венгрия",
    "location_normalized": "Венгрия",
    "brief_context": "В декабре 1848 - январе 1849 гг. австрийская армия под командованием фельдмаршала Виндишгреца вела успешное наступление против венгерских революционных сил, заняв 5 января 1849 года столицу Венгрии Пешт.",
    "information_source": "петербургские газеты",
    "information_source_type": "Официальные источники (газеты, манифесты)",
    "confidence": "High",
    "classification_confidence": "High",
    "keywords": [
```

```

        "австрийцы",
        "Венгрия",
        "газеты",
        "военные действия",
        "успехи"
    ],
    "text_fragment": "Теперь читал петербургские газеты, иностранные происшествия. Везет австрийцам, славно управляют в Венгрии, да что еще из этого всего будет?"
},
{
    "entry_id": 7,
    "event_id": "AUTHOR_PERCEPTION_OPINION_AUTH",
    "event_name": "Выражение поддержки/осуждения действий властей/конкретных государств",
    "event_subtype_custom": "Поддержка действий Австрии в Венгрии",
    "description": "Автор прямо выражает свое личное одобрение и поддержку действий Австрийской империи против венгерских повстанцев, используя оценочные суждения «Везет австрийцам, славно управляют».",
    "date_in_text": null,
    "source_date": "1849-01-07",
    "location": "Венгрия",
    "location_normalized": "Венгрия",
    "brief_context": "Официальная позиция Российской империи и консервативных кругов российского общества заключалась в поддержке монархических режимов Европы и осуждении революционных движений, в том числе венгерского.",
    "information_source": "Личное мнение автора",
    "information_source_type": "Личные наблюдения и опыт автора",
    "confidence": "High",
    "classification_confidence": "High",
    "keywords": [
        "поддержка",
        "одобрение",
        "австрийцы",
        "личное мнение",
        "Венгрия"
    ],
    "text_fragment": "Везет австрийцам, славно управляют в Венгрии..."
},
{
    "entry_id": 7,
    "event_id": "REV1848_EUR_CONTEXT_GEN",
    "event_name": "Общие упоминания Весны народов, беспорядков на Западе",
    "event_subtype_custom": "Восприятие событий как хаоса",
    "description": "Автор обобщенно характеризует всю совокупность революционных событий в Европе («на Западе») как «ералаш ужаснейший», что отражает его восприятие происходящего как всеобщего хаоса и беспорядка.",
    "date_in_text": null,
    "source_date": "1849-01-07",
    "location": "Запад",
    "location_normalized": "Европа",
    "brief_context": "К началу 1849 года, несмотря на успехи контрреволюции в отдельных регионах, общая ситуация в Европе оставалась крайне нестабильной. Продолжались военные действия в Венгрии и Италии, а политическая борьба в Германии и Франции была далека от завершения.",
    "information_source": "петербургские газеты",
    "information_source_type": "Официальные источники (газеты, манифесты)",
    "confidence": "High",
    "classification_confidence": "High",
    "keywords": [
        "Запад",
        "Европа",
        "ералаш",
        "хаос",
        "беспорядки"
    ],
    "text_fragment": "Ералаш ужаснейший, кажется, еще и теперь на Западе."
},

```

```
{
  "entry_id": 7,
  "event_id": "AUTHOR_PERCEPTION_OPINION_REFL",
  "event_name": "Размышления о политике, войне, обществе, будущем",
  "event_subtype_custom": "Тревога о будущем Европы",
  "description": "Автор выражает личную тревогу и неуверенность в будущем,
размышляя о последствиях европейских событий. Его риторический вопрос «что еще из
этого всего будет?» и оценка ситуации как «ералаш ужаснейший» свидетельствуют о
беспокойстве и негативном прогнозе.",
  "date_in_text": null,
  "source_date": "1849-01-07",
  "location": "Запад",
  "location_normalized": "Европа",
  "brief_context": "Неопределенность исхода революций 1848-1849 гг. была
характерной чертой восприятия событий современниками, так как исход многих конфликтов
и политических преобразований был еще не ясен.",
  "information_source": "Личное мнение автора",
  "information_source_type": "Личные наблюдения и опыт автора",
  "confidence": "High",
  "classification_confidence": "High",
  "keywords": [
    "размышления",
    "тревога",
    "будущее",
    "ералаш",
    "Запад"
  ],
  "text_fragment": "да что еще из этого всего будет? Ералаш ужаснейший, кажется,
еще и теперь на Западе."
}
```