# RAPID: Эффективная генерация длинного текста с использованием дополненной информации с планированием написания и обнаружением информации

Дата: 2025-03-02 00:00:00

Ссылка на исследование: https://arxiv.org/pdf/2503.00751

Рейтинг: 75
Адаптивность: 85

## Ключевые выводы:

Исследование представляет RAPID - эффективный фреймворк для генерации длинных текстов с использованием LLM, который решает проблемы галлюцинаций, тематической согласованности и задержек при создании энциклопедических статей. Основной результат - значительное превосходство RAPID над существующими методами по широкому спектру метрик оценки.

## Объяснение метода:

RAPID предлагает трехэтапный подход к созданию длинных текстов (план → поиск → написание с учетом зависимостей), который значительно повышает качество контента. Ключевые концепции (атрибутно-ориентированный поиск, последовательность написания на основе зависимостей между разделами) легко адаптируются для использования в обычных чатах, хотя полная реализация некоторых технических аспектов может быть затруднительна для неподготовленных пользователей.

Ключевые аспекты исследования: 1. **Структура RAPID** - исследование представляет фреймворк для создания длинных информационно-насыщенных текстов, состоящий из трех основных модулей: генерация плана на основе поиска, ограниченный атрибутами поиск информации и создание текста на основе плана.

**Поисково-дополненная генерация планов** - метод использует корпус примеров планов (около 2,6 миллиона) из Википедии, уточняет запрос через веб-поиск и использует релевантные примеры для создания качественного плана текста.

**Атрибутно-ограниченный поиск** - система извлекает атрибуты из плана, преобразует их в поисковые запросы и использует параллельный поиск для сбора информации, которая затем используется для уточнения плана и создания текста.

**План-ориентированная генерация текста** - создание топологического графа зависимостей между разделами для определения последовательности написания, что улучшает связность и логическую структуру текста.

**Эмпирические результаты** - эксперименты показывают, что RAPID превосходит существующие методы по качеству плана, фактической точности, связности и эффективности при создании энциклопедических статей.

Дополнение:

## # Применимость методов RAPID в стандартном чате

Методы RAPID не требуют дообучения или специального API для основных концептуальных элементов. Хотя исследователи использовали расширенные инструменты (плотные ретриверы, корпус планов), ключевые подходы можно адаптировать для стандартного чата LLM:

**Трехэтапная структура создания контента**: Пользователь может последовательно проходить этапы планирования, поиска информации и написания в обычном чате Результат: более структурированный и логичный текст

**Атрибутно-ориентированный поиск**:

Запрос к LLM: "Выдели ключевые атрибуты/понятия из темы X для поиска информации" Использование этих атрибутов для самостоятельного поиска Результат: более целенаправленный сбор информации

**План-ориентированная генерация**:

Запрос к LLM: "Определи логические зависимости между разделами плана и оптимальную последовательность написания" Следование предложенной последовательности Результат: повышение связности и логичности текста

**Итеративное уточнение плана**:

Корректировка плана на основе найденной информации Результат: лучшее соответствие плана доступным данным Эти концепции применимы в стандартном чате без специальных инструментов и значительно повышают качество длинных текстов по сравнению с прямой генерацией.

Анализ практической применимости: 1. **Структура RAPID - Прямая применимость**: Высокая. Разделение процесса создания длинных текстов на этапы (план, поиск информации, написание) можно адаптировать в обычных чатах с LLM, где пользователи могут следовать этой структуре для создания качественного контента. - **Концептуальная ценность**: Очень высокая. Понимание важности предварительного планирования и структурированного поиска информации перед

написанием помогает пользователям осознать, как улучшить взаимодействие с LLM для получения более качественных результатов. - **Потенциал для адаптации**: Высокий. Хотя фреймворк использует специализированные компоненты, основной принцип "планирование → поиск → написание с учетом зависимостей" легко адаптируется к обычным чатам.

**Поисково-дополненная генерация планов Прямая применимость**: Средняя. Обычные пользователи не имеют доступа к корпусу из 2,6 миллионов планов, но могут использовать веб-поиск для нахождения примеров планов статей по похожим темам перед созданием собственного плана. **Концептуальная ценность**: Высокая. Понимание важности изучения структуры похожих текстов перед созданием собственного плана весьма ценно для повышения качества взаимодействия с LLM. **Потенциал для адаптации**: Средний. Пользователи могут адаптировать этот подход, запрашивая у LLM создание плана на основе нескольких примеров структур, найденных в интернете.

**Атрибутно-ограниченный поиск**

**Прямая применимость**: Высокая. Метод выделения ключевых атрибутов (концепций) из плана текста и использование их для целенаправленного поиска информации может непосредственно применяться пользователями при работе с LLM. **Концептуальная ценность**: Высокая. Понимание того, как разбить сложную тему на атрибуты для более эффективного поиска информации, помогает пользователям структурировать свои запросы к LLM. **Потенциал для адаптации**: Очень высокий. Пользователи могут легко адаптировать этот подход, запрашивая у LLM выделение ключевых атрибутов из темы и затем используя их для поиска информации.

**План-ориентированная генерация текста**

**Прямая применимость**: Средняя. Создание графа зависимостей между разделами сложно реализовать напрямую, но понимание логических связей между разделами доступно обычным пользователям. **Концептуальная ценность**: Очень высокая. Осознание важности последовательности написания разделов для обеспечения связности текста крайне полезно для взаимодействия с LLM при создании длинных текстов. **Потенциал для адаптации**: Высокий. Пользователи могут запросить у LLM определить логические зависимости между разделами плана и следовать рекомендованной последовательности написания.

**Эмпирические результаты**

**Прямая применимость**: Низкая. Конкретные метрики и результаты экспериментов имеют ограниченную практическую ценность для обычных пользователей. **Концептуальная ценность**: Средняя. Понимание того, какие аспекты влияют на качество генерируемого текста (фактическая точность, связность, информативность), помогает пользователям формулировать более эффективные запросы. **Потенциал для адаптации**: Низкий. Методология оценки сложно адаптируема для использования обычными пользователями. Сводная оценка

полезности: На основе анализа определяю общую оценку полезности исследования для широкой аудитории: **78**.

Исследование RAPID предлагает исключительно ценную методологию для создания качественных длинных текстов, которая может быть адаптирована обычными пользователями LLM. Основные концепции (предварительное планирование, атрибутно-ориентированный поиск информации, последовательность написания с учетом зависимостей между разделами) представляют высокую практическую ценность и могут быть реализованы в обычных чатах без специализированных инструментов.

**Контраргументы к оценке:**

**Почему оценка могла бы быть выше:** Исследование предлагает четкую и логичную структуру процесса, которая может существенно улучшить качество длинных текстов, создаваемых с помощью LLM. Принципы легко понимаемы и могут быть применены даже неподготовленными пользователями.

**Почему оценка могла бы быть ниже:** Полная реализация метода требует доступа к специализированным инструментам (корпус планов, плотные ретриверы, параллельные поисковые запросы), которые недоступны обычным пользователям. Также создание графа зависимостей между разделами может быть сложным для неподготовленных пользователей.

После рассмотрения контраргументов, корректирую оценку до **75**. Хотя методология исключительно ценна, некоторые аспекты требуют адаптации для широкой аудитории.

Оценка **75** отражает: 1. Высокую практическую ценность трехэтапного подхода к созданию длинных текстов 2. Возможность адаптации основных концепций для использования в обычных чатах 3. Значительное улучшение качества генерируемого контента при применении принципов 4. Необходимость определенной адаптации технических аспектов для широкой аудитории 5. Универсальность подхода для различных типов длинных информационно-насыщенных текстов

Уверенность в оценке: Моя уверенность в оценке: **очень сильная**.

Исследование представляет четкую методологию с понятными компонентами, которые могут быть адаптированы пользователями разного уровня подготовки. Эмпирические результаты убедительно демонстрируют эффективность подхода, а человеческая оценка подтверждает преимущества метода. Структура RAPID логична и соответствует естественному процессу создания качественного контента, что делает ее интуитивно понятной и применимой.

Оценка адаптивности: Оценка адаптивности: **85**.

RAPID демонстрирует высокий потенциал для адаптации по следующим причинам:

1) Трехэтапная структура (планирование, поиск информации, написание с учетом зависимостей) может быть непосредственно применена пользователями в обычном чате с LLM путем последовательного выполнения этих этапов.

2) Концепция выделения ключевых атрибутов из плана для целенаправленного поиска информации легко реализуема в обычных чатах – пользователи могут запросить у LLM выделить ключевые понятия из темы и использовать их для поиска.

3) Идея создания логической последовательности написания разделов на основе их зависимостей может быть адаптирована путем запроса у LLM определить оптимальный порядок написания разделов плана.

4) Принцип использования примеров структур похожих текстов для создания качественного плана может быть реализован путем поиска и предоставления LLM примеров структур аналогичных текстов.

Основные концепции исследования могут быть абстрагированы до общих принципов взаимодействия с LLM: структурированный подход к созданию контента, важность предварительного планирования, целенаправленный сбор информации и логическая последовательность написания. Эти принципы универсальны и могут применяться для различных задач создания длинных текстов.

|| <Оценка: 75> || <Объяснение: RAPID предлагает трехэтапный подход к созданию длинных текстов (план → поиск → написание с учетом зависимостей), который значительно повышает качество контента. Ключевые концепции (атрибутно-ориентированный поиск, последовательность написания на основе зависимостей между разделами) легко адаптируются для использования в обычных чатах, хотя полная реализация некоторых технических аспектов может быть затруднительна для неподготовленных пользователей.> || <Адаптивность: 85>

## Prompt:

Применение методологии RAPID в промптах для GPT
**Ключевые принципы RAPID для использования в промптах**

Исследование RAPID предлагает трехкомпонентный подход, который можно эффективно адаптировать для работы с GPT:

**Предварительное планирование с поиском Атрибутно-ориентированный сбор информации Структурированная генерация на основе плана**
**Пример промпта на основе методологии RAPID**

[======]

Задача: Создание энциклопедической статьи о [ТЕМА]

### Этап 1: Уточнение темы и планирование

Перед тем как составить план статьи, уточни ключевые аспекты темы [ТЕМА]. Определи: - О каком именно [ТЕМА] идет речь (избегай неоднозначностей) - Какие основные категории информации должны быть включены - Какая структура будет наиболее подходящей для данной темы

### Этап 2: Создание структурированного плана

На основе уточненной информации создай детальный план статьи, включающий: - Введение с кратким определением [ТЕМА] - 4-6 основных разделов с подразделами - Логическую последовательность разделов, учитывающую зависимости между темами

### Этап 3: Определение ключевых атрибутов для каждого раздела

Для каждого раздела плана определи 3-5 ключевых атрибутов или вопросов, на которые нужно ответить. Например: - Раздел "История": происхождение, ключевые даты, этапы развития, значимые события - Раздел "Характеристики": технические параметры, особенности, сравнение с аналогами

### Этап 4: Генерация содержания по плану

Теперь, используя созданный план и определенные атрибуты, напиши полную статью, соблюдая: - Логическую связность между разделами - Полноту раскрытия каждого атрибута - Фактическую точность информации - Энциклопедический стиль изложения [======]

### Почему это работает

Данный промпт использует ключевые принципы RAPID:

**Предотвращение галлюцинаций** через предварительное уточнение темы и планирование **Структурированный подход** через создание детального плана с логическими связями **Атрибутно-ориентированный сбор информации** через определение ключевых атрибутов для каждого раздела **Повышение связности** через генерацию контента на основе структурированного плана Такой подход позволяет получить более качественный, структурированный и фактически точный результат при работе с GPT, особенно при создании длинных информационных текстов.