На сайте Банка России публикуются исследовательские работы сотрудников. Стороннему пользователю может быть затруднительно их искать. Поэтому есть потребность упростить поиск данных работ. Айтишники нам сделают на сайте теги по темам. Например, «инфляция», «модельный аппарат»,… Таких тегов будет примерно 20 штук. Каждая статья будет принадлежать 1-5 тегам (пока условно). При нажатии на конкретный тег должны будут появляться результаты со статьями по данной теме.

Наша задача заключается в том, чтобы придумать эти 20 тегов. Согласовать с руководством финальный список тегов, а далее каждому препринту присвоить 1-5 тегов и передать этот список айтишникам.

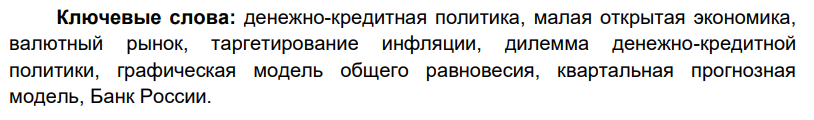
Сами исследовательские работы:

1. В разделе Исследования есть подраздел с препринтами сотрудников Банка России «Доклады об экономических исследованиях» <https://www.cbr.ru/ec_research/ser/>
2. В разделе Денежно-кредитная политика есть подраздел с записками по модельному аппарату <https://www.cbr.ru/dkp/system_p/>

На первом этапе необходимо подготовить список тегов. Каждый тег должен быть не длиннее 3-х слов. Точно будут теги

1. «инфляция»,
2. «денежно-кредитная политика»,
3. «макропруденциальная политика»,
4. «прогноз»,
5. «финансовый рынок»,
6. «валютный курс»,
7. «модельный аппарат»…

Остальные нужно еще придумать. Для этого предлагаем отталкиваться от ключевых слов в препринтах.

В каждом препринте есть раздел «ключевые слова» или “key words”. Типа такого:

Нужно выписать все эти ключевые слова из всех препринтов (препринтов порядка штук 150 или больше) и сгруппировать их по темам (на примере данного скрина: «квартальная прогнозная модель» - это тема «модельный аппарат» и «прогноз»). Наиболее часто встречающиеся (темы) и станут тегами.

На втором этапе согласовать список тегов с Вадимом/Аленой.

На третьем этапе каждому из препринтов присвоить 1-5 тегов (примерно). В результате 3-го этапа должна получиться таблица в экселе. Первый столбец – название препринта, второй – ссылка на него с сайта БР, третий и далее – теги, которые соответствуют конкретному препринту.

Все это можно сделать вручную или прибегая к любым доступным способам.

**Промпт для ИИ:**

Привет. Я сейчас работаю над следующим заданием для ЦБ РФ. На сайте Банка России публикуются исследовательские работы сотрудников. Стороннему пользователю может быть затруднительно их искать. Поэтому есть потребность упростить поиск данных работ. Айтишники нам сделают на сайте теги по темам. Например, «инфляция», «модельный аппарат»,… Таких тегов будет примерно 20 штук. Каждая статья будет принадлежать 1-5 тегам (пока условно). При нажатии на конкретный тег должны будут появляться результаты со статьями по данной теме. Наша задача заключается в том, чтобы придумать эти 20 тегов. Согласовать с руководством финальный список тегов, а далее каждому препринту присвоить 1-5 тегов и передать этот список айтишникам. 7 тегов точно должны быть такие: «инфляция», «денежно-кредитная политика»,«макропруденциальная политика», «прогноз», «финансовый рынок», «валютный курс», «модельный аппарат». Осталось придумать 13 тегов. Я уже смог написать код, который выдает для каждой статьи ключевые слова. Сможешь на их основе (согласно частоте их использования), придумать названия для тегов? Я еще не считал частоту использования того или иного тега. Это предстоит также сделать тебе.

Я могу прислать набор ключевых слов в виде таблицы или в текстовом формате. Например, могу прислать таблицу Excel с такими столбцами: title, keyword\_1, keyword\_2, keyword\_3... Что думаешь?

*Отправляю файл эксель со столбцом «title» и «keywords\_1», «keywords\_2»…*

Вот таблица. Только учти, что в некоторых статьях нет ключевых слов (в таком случае в таблице в столбце "keyword\_1" будет написано "Нет ключевых слов"). В таких статьях ориентируйся по их названию (какие ключевые слова могли бы быть, исходя из названия). В других случаях не учитывай названия статей (первый столбец) при подсчете частотности тех или иных ключевых слов.