

Сели мы писать парсер на Amazon и столкнулись с 2 проблемами:

1 – Для российский IP Amazon выдаёт очень ограниченное кол-то товаров.

2 – Арендовали хостинг в США с американскими IP сразу капча. Прокси не помогли (либо их нужно очень много).

Короче тут наши попытки и закончились.

Ищем подрядчика кто сам выберет хостинг, IP и напишет парсер.

Возможно обойти капчу и сделать все по-болеому вам поможет официальный API Amazon <https://aws.amazon.com/ru/api-gateway/pricing/>

Если там нужно что-то заплатить Амазону за наши запросы к нужно для нас категории (/alexa-skills) мы заплатим, там недорого вроде.

ТЗ на парсер

Заходим по этой ссылке <https://www.amazon.com/alexa-skills/b?ie=UTF8&node=13727921011>

Слева видим разделы, которые нам нужны:

- Business & Finance
- Communication
- Connected Car
- Education & Reference
- Enterprise
- Food & Drink
- Games & Trivia
- Health & Fitness
- Home Services
- Kids
- Lifestyle
- Local
- Movies & TV
- Music & Audio
- News
- Novelty & Humor
- Productivity
- Shopping
- Smart Home
- Social
- Sports
- Travel & Transportation
- Utilities
- Weather

Нужно парсить все. В этих разделах вот столько страниц (такое количество с американских IP или когда вы выставляете в браузере адрес США). Для России страниц гораздо меньше.

Business & Finance **214**

Communication **30**

Connected Car **8**

Education & Reference больше **400** стр, больше 10.000 приложений

Enterprise **1**

Food & Drink **84**

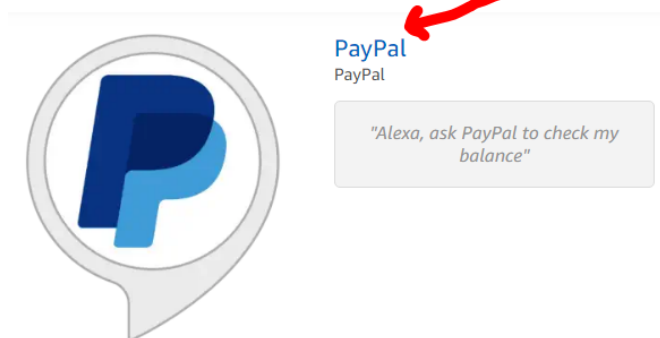
Games & Trivia больше **400**

Health & Fitness **121**
Home Services **18**
Kids **169**
Lifestyle больше **400**
Local **78**
Movies & TV **55**
Music & Audio больше **400**
News **398**
Novelty & Humor **210**
Productivity **232**
Shopping **18**
Smart Home **138**
Social **77**
Sports **94**
Travel & Transportation **73**
Utilities **51**
Weather **53**

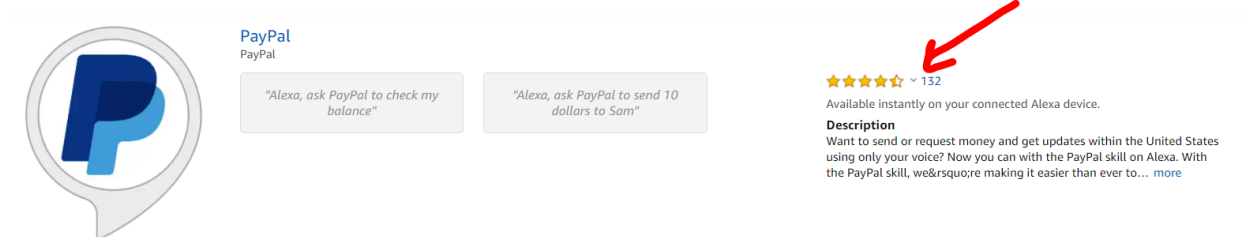
На каждой странице по 16 приложений.

Что нам нужно:

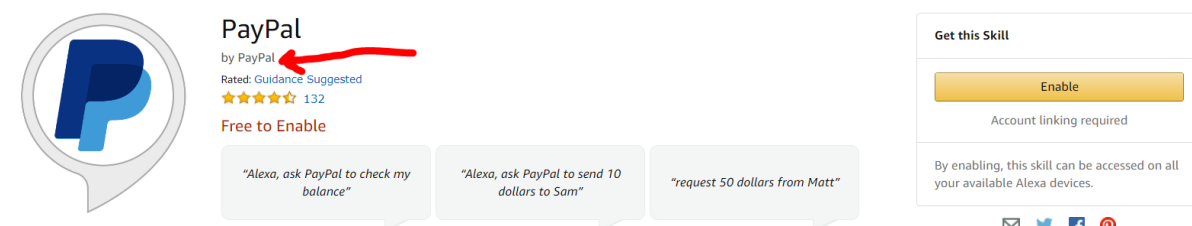
1 - Название со ссылкой на Amazon



2 – Количество оценок



3 – Разработчик (указан на внутренней странице)



4 – Дата добавления приложения в нашу таблицу

И фильтр по каждому столбцу, чтобы можно было сортировать. Получится примерно такая таблица:

Business & Finance v				
Id v	Skills v	Developer v	Ratings v	Date Added v
1	PayPal	by developer name	132	09.11.2019

ДОПОЛНИТЕЛЬНАЯ ИНФА:

1 – Нужен фильтр по категориям вверху таблицы.

2 – Export CSV

3 – Нужен фильтр по каждому столбцу, чтобы можно было сортировать.

4 – Чтобы сильно не грузить БД (т.к. там 100.000 приложений), то добавляем только приложения, у которых больше 50 оценок.

5 – Нужно сделать возможность, чтобы в будущем мы могли сами раз в месяц запускать этот парсер и по дате видеть кто из новых добавился (т.е. перешел отметку в 50 оценок).

6 – Не гонитесь за скоростью, чтобы не получить бан или капчу. Даже если он будет парсить 2 дня – ничего страшного, мы не торопимся.

Дизайн не нужен, вот такое пойдет:

Business & Finance v			
Export to CSV			
id ↕	name	total ↕	dt ↕
263	Make Me Smart	2131	2019-11-07
262	Mastermind - AI Assistant for Software Dev Teams	1540	2019-11-07
265	Coinbase Bitcoin/Ethereum Price	148	2019-11-07
260	Thirty Second Money Tip with Nick True	134	2019-11-07
267	McDonald's Apply Thru	127	2019-11-07
257	The Pitch with Amy Summers	116	2019-11-07
258	Front Row Entrepreneur	102	2019-11-07
264	YNAB	76	2019-11-07
261	Run To Your Leadership Challenges!	59	2019-11-07
259	Pete's Points - Business and Leadership Strategies	50	2019-11-07
266	PSE&G (Public Service Electric & Gas)	50	2019-11-07