

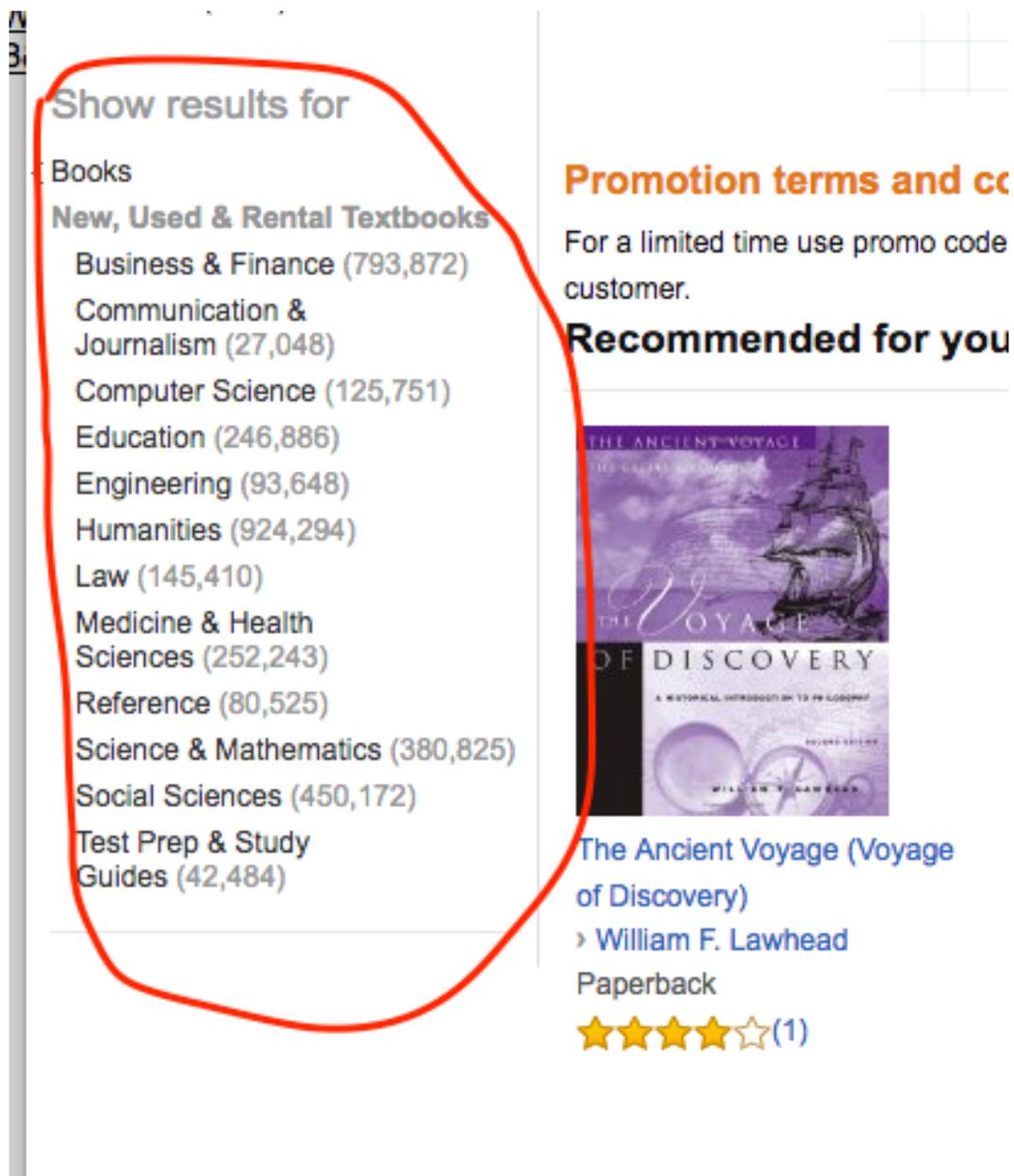
Парсим Textbooks с [amazon.com](https://www.amazon.com)

## 1) Собрать список ASIN

нужно распарсить раздел

[https://www.amazon.com/New-Used-Textbooks-Books/b/ref=nav\\_shopall\\_tb?ie=UTF8&node=465600](https://www.amazon.com/New-Used-Textbooks-Books/b/ref=nav_shopall_tb?ie=UTF8&node=465600)

Перебирать нужно вот эти категории (внизу в меню)



The screenshot shows a portion of an Amazon website. On the left, a list of textbook categories is displayed under the heading "Show results for Books". A red hand-drawn circle highlights this list. To the right, there is a promotional banner for "Promotion terms and conditions" and a "Recommended for you" section. The recommended book is "The Ancient Voyage (Voyage of Discovery)" by William F. Lawhead, a paperback with a 4.5-star rating based on one review.

**Show results for**

Books

**New, Used & Rental Textbooks**

- Business & Finance (793,872)
- Communication & Journalism (27,048)
- Computer Science (125,751)
- Education (246,886)
- Engineering (93,648)
- Humanities (924,294)
- Law (145,410)
- Medicine & Health Sciences (252,243)
- Reference (80,525)
- Science & Mathematics (380,825)
- Social Sciences (450,172)
- Test Prep & Study Guides (42,484)

**Promotion terms and conditions**

For a limited time use promo code customer.

**Recommended for you**

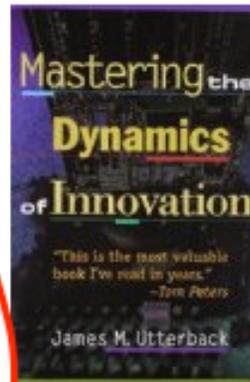
**The Ancient Voyage (Voyage of Discovery)**  
› William F. Lawhead  
Paperback  
★★★★☆ (1)

В каждой категории есть своя пачка под категорий

Show results for

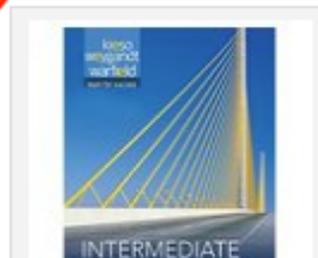
- < Books
- < New, Used & Rental Textbooks
- Business & Finance**
  - Accounting (22,139)
  - Banking (5,564)
  - Business Communication (4,028)
  - Business Development (23,891)
  - Business Ethics (3,366)
  - Business Law (12,526)
  - Economics (550,195)
  - Entrepreneurship (6,354)
  - Finance (15,954)
  - Human Resources (10,592)
  - International Business (39,481)
  - Investments & Securities (7,026)
  - Management (39,818)
  - Marketing (12,010)
  - Real Estate (3,823)
  - Sales (2,090)

Recommended for you



Mastering the Dynamics of Innovation  
James M. Utterback  
Paperback  
★★★★☆ (16)  
\$40.00 \$33.32 Prime  
Why recommended?

Shop by category



например:

Категорий дальше Зего уровня вроде бы нет.

Show results for

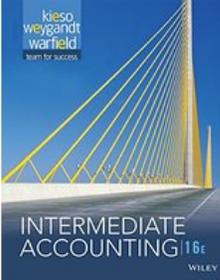
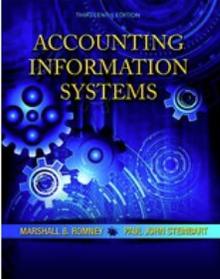
- < Books
- < New, Used & Rental Textbooks
- < Business & Finance
- Accounting**

R

смотрим пример выдачи:

[https://www.amazon.com/s/ref=lp\\_491564\\_il\\_to\\_stripbooks?rh=n%3A283155%2Cn%3A%212349030011%2Cn%3A465600%2Cn%3A468220%2Cn%3A491564&ie=UTF8&qid=1472525754&lo=none](https://www.amazon.com/s/ref=lp_491564_il_to_stripbooks?rh=n%3A283155%2Cn%3A%212349030011%2Cn%3A465600%2Cn%3A468220%2Cn%3A491564&ie=UTF8&qid=1472525754&lo=none)

---

	<p><b>Intermediate Accounting</b> Apr 4, 2016 by Donald E. Kieso and Jerry J. Weygandt</p> <p>Hardcover <b>\$315.15</b> ✓Prime FREE Shipping on eligible orders More Buying Choices <b>\$289.92</b> used &amp; new (42 offers)</p> <hr/> <p>Kindle Edition from <b>\$79.09</b> to rent <b>\$108.50</b> to buy Auto-delivered wirelessly</p> <hr/> <p>Other Formats: <a href="#">Loose Leaf</a></p>	★★★★
	<p><b>Best Seller</b></p> <p><b>Accounting Information Systems (13th Edition)</b> Jan 19, 2014 by Marshall B. Romney and Paul J. Steinbart</p> <p>Hardcover <b>\$83.07 - \$234.90</b> to rent ✓Prime <b>\$274.74</b> to buy ✓Prime FREE Shipping on eligible orders Only 20 left in stock - order soon. More Buying Choices <b>\$141.75</b> used &amp; new (96 offers)</p> <hr/> <p>Kindle Edition from <b>\$131.99</b> to rent <b>\$192.99</b> to buy Auto-delivered wirelessly</p> <hr/> <p>Other Formats: <a href="#">Paperback</a></p>	★★★★ <a href="#">Trade in</a>
	<p><b>Cost Accounting (15th Edition)</b> Feb 14, 2014 by Charles T. Horngren and Srikant M. Datar</p> <p>Hardcover</p>	★★★★

Под названием книги лежит ссылка вида

[https://www.amazon.com/Intermediate-Accounting-Donald-E-Kieso/dp/1118743202/ref=lp\\_491564\\_1\\_1?s=books&ie=UTF8&qid=1472525761&sr=1-1](https://www.amazon.com/Intermediate-Accounting-Donald-E-Kieso/dp/1118743202/ref=lp_491564_1_1?s=books&ie=UTF8&qid=1472525761&sr=1-1)

Из всех ссылок нужно извлечь набор цифр между dp/ и /ref (могут быть и буквы и цифры):

[https://www.amazon.com/Intermediate-Accounting-Donald-E-Kieso/dp/1118743202/ref=lp\\_491564\\_1\\_1?s=books&ie=UTF8&qid=1472525761&sr=1-1](https://www.amazon.com/Intermediate-Accounting-Donald-E-Kieso/dp/1118743202/ref=lp_491564_1_1?s=books&ie=UTF8&qid=1472525761&sr=1-1)

Перебираем все страницы в пагинаторе и собираем эти цифры из всех ссылок что представлены на страницах.

## 2) Для базы ASIN собираем цены

Когда у нас готова база ASIN

дергаем URL такого вида

[https://www.amazon.com/gp/offer-listing/1118743202/ref=olp\\_f\\_new?ie=UTF8&f\\_new=true&f\\_used=true&f\\_usedAcceptable=true&f\\_usedGood=true&f\\_usedLikeNew=true&f\\_usedVeryGood=true&shipPromoFilter=1](https://www.amazon.com/gp/offer-listing/1118743202/ref=olp_f_new?ie=UTF8&f_new=true&f_used=true&f_usedAcceptable=true&f_usedGood=true&f_usedLikeNew=true&f_usedVeryGood=true&shipPromoFilter=1)

где вместо **1118743202** подставляются цифры/буквы ASIN которые мы собрали до этого.

По ссылке видим такую картину(на галки на скринах не смотрим):

Refine by [Clear all](#)

Shipping  
 Prime  
 Free shipping

Condition  
 New  
 Rental  
 Used  
 Like New  
 Very Good  
 Good  
 Acceptable

Price + Shipping  Condition [\(Learn more\)](#) Delivery

**\$315.15** Prime  
+ \$28.36 estimated tax

**New**

- Free Two-day Shipping
- [Shipping rates and return policy.](#)

ИЛИ

Refine by [Clear all](#)

Shipping  
 Prime  
 Free shipping

Condition  
 New  
 Rental  
 Used  
 Like New  
 Very Good  
 Good  
 Acceptable

Price + Shipping  Condition [\(Learn more\)](#) Delivery

<b>\$162.52</b> Prime + \$14.63 estimated tax	<b>Used - Good</b> Ships direct from Amazon! Qualifies for Prime Shipping and FREE s... » <a href="#">Read more</a>	<input type="text" value="FULFILLMENT BY AMAZON"/>
<b>\$168.09</b> Prime + \$15.13 estimated tax	<b>Used - Acceptable</b> Ships quick from Amazon! Qualifies for Prime Shipping and FREE st... » <a href="#">Read more</a>	<input type="text" value="FULFILLMENT BY AMAZON"/>
<b>\$169.95</b> Prime + \$15.30 estimated tax	<b>New</b> New, book only does not contain access card or code.	<input type="text" value="FULFILLMENT BY AMAZON"/>
<b>\$206.38</b> Prime + \$18.57 estimated tax	<b>New</b>	<input type="text" value="FULFILLMENT BY AMAZON"/>

- Free Two-Day Shipping: Get it Wednesday, A order within **3hr 52min** )
- [Shipping rates and return policy.](#)
- Free Two-day Shipping
- [Shipping rates and return policy.](#)
- Free Two-Day Shipping: Get it Thursday, Sep order within **20hr 53min** )
- [Shipping rates and return policy.](#)
- Free Two-day Shipping
- [Shipping rates and return policy.](#)

Или

Refine by [Clear all](#)

**Shipping**

- Prime
- Free shipping

**Condition**

- New
- Rental

There are currently no listings for this search. Try a different refinement.

Результаты складываем в mysql таким образом  
asin | first price | first condition | price array

где

asin = asin

first price - первая цена в таблице

first condition - condition для первого элемента в таблице

price array - цены через запятую

пример

**1118743202 | 315.15 | new | 315.15**

**0077862279 | 162.52 | Used -Good | 162.52, 168.09, 169.95,  
206.308**

**если данных нету - пишем NULL**