

استخراج اطلاعات مورد نیاز از سایت imdb

با توجه به بخش توضیحات کلی تمرین برای انجام مرحله اول و استخراج اطلاعات از این [لینک](#) استفاده کنید. این کار را با استفاده از روش‌های مختلف کراولینگ این اطلاعات انجام دهید. هر یک از این 250 فیلم خود شامل لینکی هستند که شما را به صفحه آن فیلم هدایت می‌کند.

پس از ورود به صفحه هر فیلم شما باید اطلاعات زیر را استخراج کنید:

۱. نام یا عنوان فیلم (title)
۲. سال انتشار فیلم (year)
۳. درجه سنی فیلم (parental_guide)
۴. مدت زمان فیلم بر حسب دقیقه (runtime)
۵. ژانر یا ژانرهای فیلم (genre)
۶. نام کامل کارگردان یا کارگردانان (directore)
۷. نام کامل نویسنده یا نویسندگان (writer)
۸. نام کامل بازیگر یا بازیگران (star)
۹. توضیحات فیلم یا خلاصه فیلم (storyline) (اختیاری)
۱۰. میزان فروش فیلم در منطقه آمریکا و کانادا (gross_us_canada)

▼ راهنمایی مربوط به عنوان فیلم

دقت کنید برای اینکه اطلاعات شما برای فیلم‌هایی که به زبان انگلیسی نیستند مطابق قسمت مشخص شده در سوال باشد باید در `header` درخواست خود `accept-language` را انگلیسی تنظیم کنید.

▼ راهنمایی مربوط به توضیحات فیلم

دقت کنید که توضیحات فیلم یا `storyline` در سایت به صورت پویا یا `dynamic` می‌باشد و الزاماً روش‌هایی که اطلاعات دیگر را جمع‌آوری کرده‌اید ممکن است بدست نیاید. به عنوان مثال می‌توانید از `selenium` استفاده کنید، همچنین فقط برای `stoyline` مجاز به استفاده از `api` سایت هستید. همچنین برای راحتی شما امتیاز توضیحات فیلم به صورت اختیاری اعمال می‌شود.

▼ توجه

با توجه به این که اطلاعات مورد نیاز در بخش های مختلفی از صفحه فیلم ممکن است تکرار شود اکیدا توصیه می شود که اطلاعات را از قسمت های مشخص شده جمع اوری کنید.

برای تشخیص هر یک از موارد بالا و جایگاه آنها در صفحه فیلم می توانید از عکس های زیر استفاده کنید.

Life Is Beautiful ¹

Original title: *La vita è bella*

² 1997 ³ PG-13 ⁴ 1h 56m

⁵ Comedy Drama Romance

When an open-minded Jewish waiter and his son become victims of the Holocaust, he uses a perfect mixture of will, humor and imagination to protect his son from the dangers around their camp.

Director ⁶ Roberto Benigni

Writers ⁷ Vincenzo Cerami · Roberto Benigni

Stars ⁸ Roberto Benigni · Nicoletta Braschi · Giorgio Cantarini

IMDb Rating 8.6/10 (708K)

YOUR RATING ☆ Rate

POPULARITY 648 ↑ 38

2 VIDEOS

99+ PHOTOS

Search on Amazon search for Blu-ray and DVD

Add to Watchlist Added by 662K users

1.4K User reviews **138** Critic reviews

59 Metascore

IMDbPro See production, box office & company info

Storyline

[Edit](#)

In 1930s Italy, a carefree Jewish waiter named Guido starts a fairy tale life by courting and marrying a lovely woman from a nearby city. Guido and his wife have a son and live happily together until the occupation of Italy by German forces. In an attempt to hold his family together and help his son survive the horrors of a Jewish Concentration Camp, Guido imagines that the Holocaust is a game and that the grand prize for winning is a tank.—Anthony Hughes <husnock31@hotmail.com>

[world war two](#)
[jewish](#)
[nazi occupation](#)
[holocaust](#)
[tank](#)
[83 more](#)

9

[Plot summary](#) · [Plot synopsis](#)

Taglines

An unforgettable fable that proves love, family and imagination conquer all. (Canadian one sheet) >

Genres [Comedy](#) · [Drama](#) · [Romance](#) · [War](#)

Motion Picture Rating (MPAA) Rated PG-13 for holocaust-related thematic elements >

Parents guide >

Box office

Budget

\$20,000,000 (estimated)

Gross US & Canada

\$57,563,264

10

Opening weekend US & Canada

\$118,920 · Oct 25, 1998

Gross worldwide

\$230,099,013

IMDbPro [See detailed box office info on IMDbPro](#)

همچنین برای ذخیره اطلاعات در پایگاه داده نیاز به شناسه (id) فیلمها و افراد موجود در سایت است. برای بدست آوردن شناسه فیلمها و افراد باید به لینک هر یک توجه کرد و قسمت عددی بعد از title و name را بدون در نظر گرفتن قسمت رشته‌ای آنها (nm و tt) به عنوان شناسه لحاظ کرد. برای درک بهتر شناسه به مثالهای زیر توجه کنید:

<https://www.imdb.com/title/tt0111161/>

شناسه فیلم: 0111161

<https://www.imdb.com/name/nm0000209/>

شناسه فرد: 0000209

تذکر: شما برای استخراج اطلاعات مجاز به استفاده از api نیستید و در صورت استفاده از api سایت imdb نمره‌ای به بخش اول تعلق نخواهد گرفت.

ارسال پاسخ

پاسخ شما باید یک فایل پایتونی با نام `crawl.py` باشد که با اجرای آن مطلوبات مساله به صورت کامل انجام شود.