**1. 网址**：<https://www.ted.com/talks>

**2. 爬取目标**：

2.1 视频列表页数据：

视频标题

视频链接

演讲时长（分钟）

2.2 视频详情页数据

播放量

发布时间

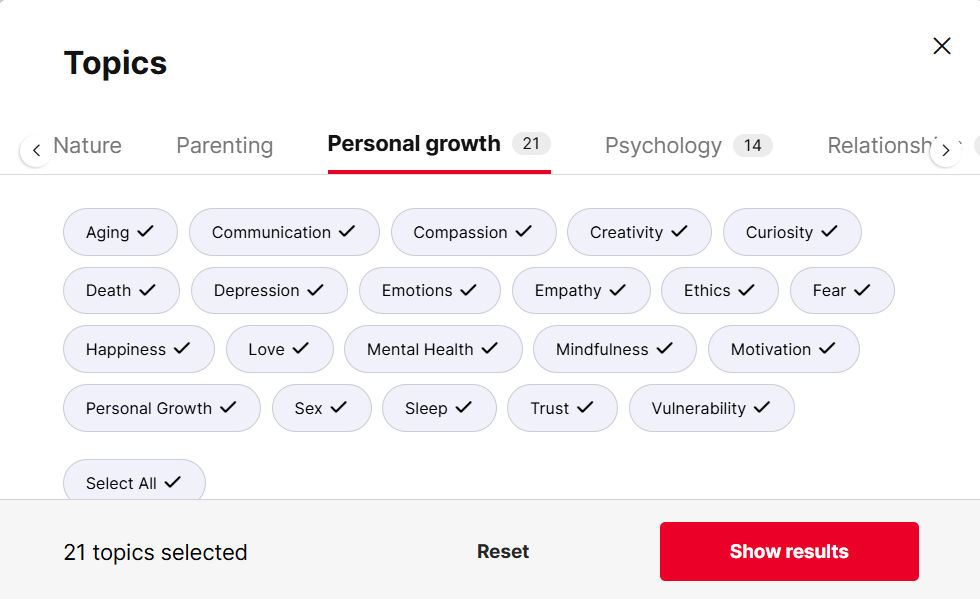
英文 transcript txt（视频右下角点击“read transcript 会在右侧显示”）

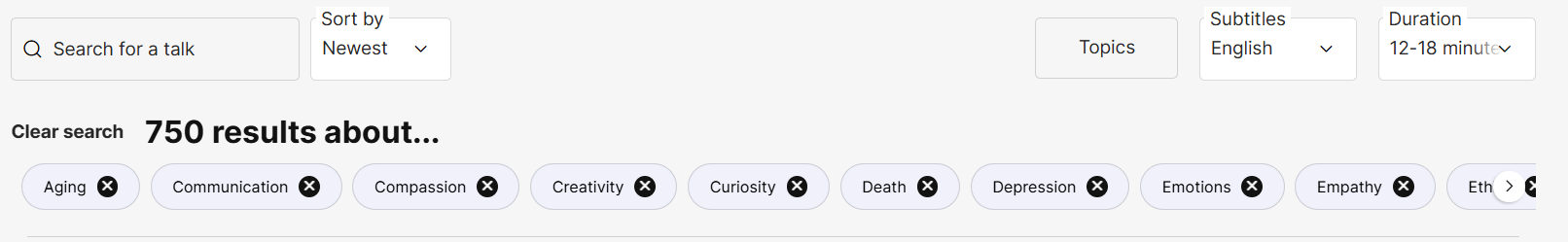
**3. 筛选条件：**

时长区间：12-18 min

Subtitles：English

Topics里勾选Personal growth下所有小分类





**4. 交付（UTF-8 ）：**

①包含所有爬取目标的csv文件（按照播放量从小到大排列），第一列添加标注001，002……750（如果真的有这么多）

②播放量前 100 的 transcript 文件，一个文件一个transcript，文件命名按如下

high\_views\_001. txt

……

high\_views\_100. Txt

③播放量最后100的transcript文件，要求同上，命名

low\_views\_650. Txt

……

high\_views\_750. txt

ps：如果没有 transcript 或 transcript 不完整，需标记