排序算法

# 通用排序算法

## 地位

随处可见，很多地方都需要用的。凡是涉及到了多个条目的展示。

可以不止是按照一个维度有数据库返回一个排序结果。最低级。

可以根据当前的场景，以及需要的效果，自己去设计一个综合考虑多个维度及目的来对一系列的数据进行排序的一个综合排序算法。

## 引入

先考虑最基本的case

1. 按时间排序 （最简单）
2. 单位时间内的交互数，del.icio.us按1小时内收藏排
3. 总交互数，回复数，view数，按总点赞数

View 100 post 10 keep 20 like 5

权重 1 5 10 10

1. 评论数加权
2. 其他

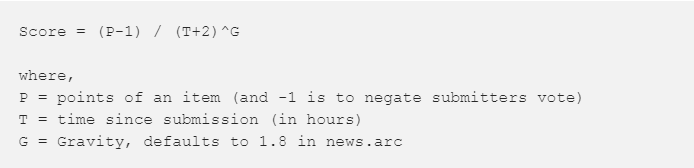
# 经典案例

## Hackernews

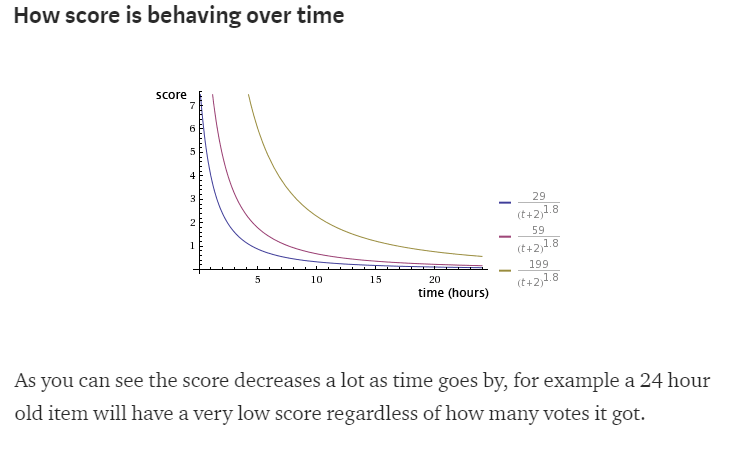
### 网站介绍

http://news.ycombinator.com/

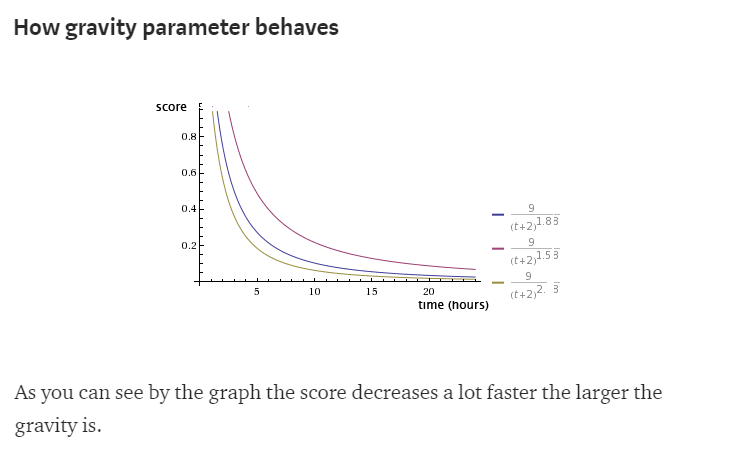
### 算法步骤



### 评分和时间对排名的影响



### 参数G对排名的影响



### 参考资料

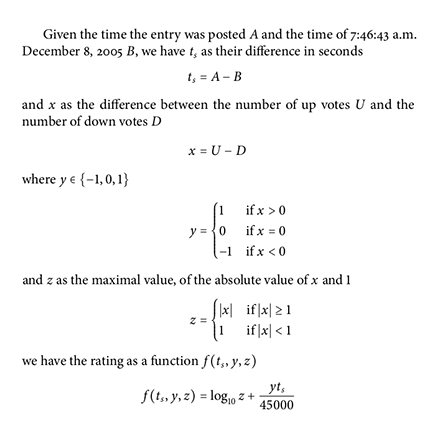
<https://medium.com/hacking-and-gonzo/how-hacker-news-ranking-algorithm-works-1d9b0cf2c08d>

## Reddit

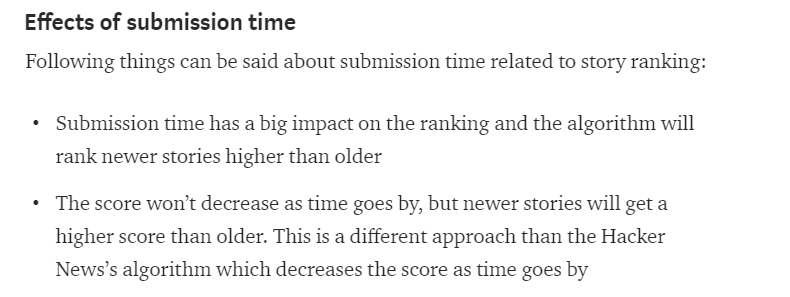
### 链接

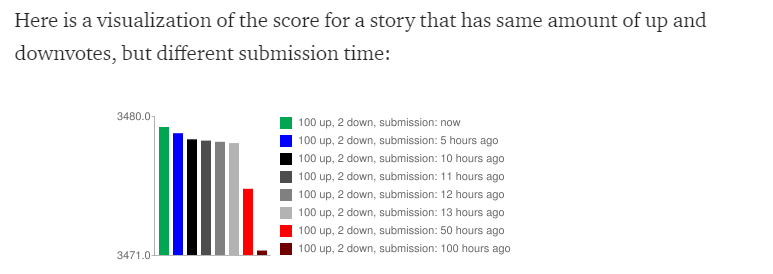
https://www.reddit.com/

### 算法步骤

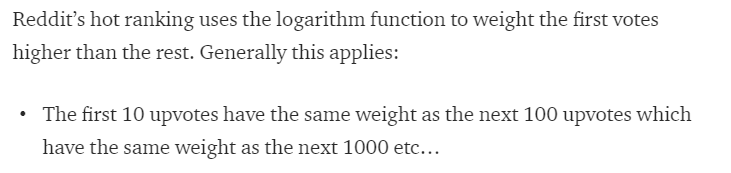
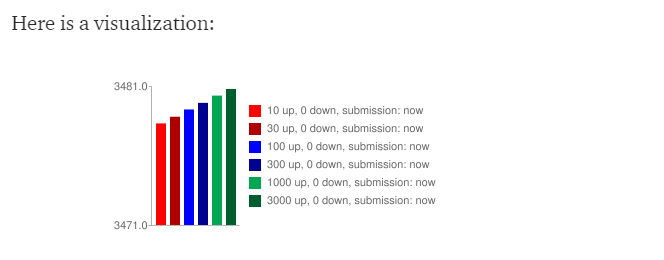


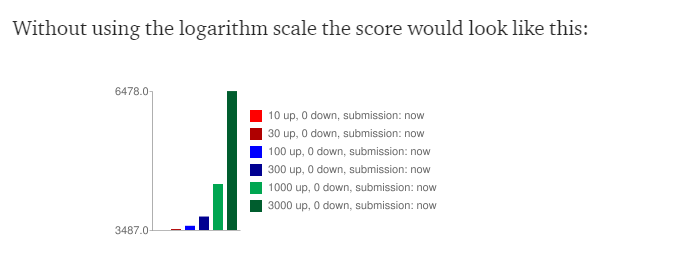
### 参数t对算法的影响





### 投票数对算法的影响



### 小结

Conclusion of Reddit’s story ranking

* Submission time is a very important parameter, generally newer stories will rank higher than older
* The first 10 upvotes count as high as the next 100. E.g. a story that has 10 upvotes and a story that has 50 upvotes will have a similar ranking
* Controversial stories that get similar amounts of upvotes and downvotes will get a low ranking compared to stories that mainly get upvotes

### 参考资料

<https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9>

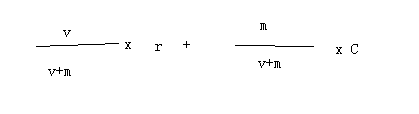
## IMDB

### 页面 Top250

IMDB <http://www.imdb.com/chart/top>

### 排序算法





R = average for the movie (mean) = (Rating) （是用普通的方法计算出的平均分）

v = number of votes for the movie = (votes) （有效投票人数note: for this top 250， only votes from regular voters are considered.）

m = minimum votes required to be listed in the top 250 (currently 1250) （进入imdb top 250需要的最小票数，只有三两个人投票的电影就算得满分也没用的）

C = the mean vote across the whole report (currently 6.9) （目前所有电影的平均得分）

### 结论

研究一下公式里的C和m可以发现，IMDB为Top 250的影片每部都增加了1250个votes，且都为6.9分（平均水准），这样可以保证每部影片都有1250+个评分，再用不停产生的votes来修正修正，v/(v+m)的值会越来越大，权重将越来越大，长期来看，该得分也就越能反映影片的真实水平。

# 扩展

关注，经常去了解一些网站里的一些算法。