

Assignment 5

CS 734: Introduction to Information Retrieval

Fall 2017

Grant Atkins

Finished on December 15, 2017

1

Question

Answer

2

Question

Answer

3

Question

10.6. Find two examples of document filtering systems on the Web. How do they build a profile for your information need? Is the system static or adaptive?

Answer

Two examples of document filtering systems on the web are Amazon and Medium. When buying tech accessories such as HDMI cables, keyboards, or mouses I usually go to Amazon. Amazon keeps track of the things I browse as well as what I previously bought and attempts to make recommendations as shown in Figure 1. This kind of filtering is an adaptive system. If I start looking at a specific book frequently, Amazon will eventually start recommending the same book if I haven't bought it, or books that are similar to the one I viewed.



Figure 1: Amazon recommended buys

Another example of a document filtering system is Medium, a popular article sharing website much like blogspot. I often enjoy reading tech related articles for machine learning, javascript, and much more on Medium. Medium keeps track of users viewing history and then when a user visits the home page of the website it offers them lists of articles to read. It has a “You might like” section where for each article it even says “Based on your interests.” This shows that it has filtering in place that is adaptive to my viewing history and tries to find the best trending/matching documents for my reading. My results are shown in Figure 2.

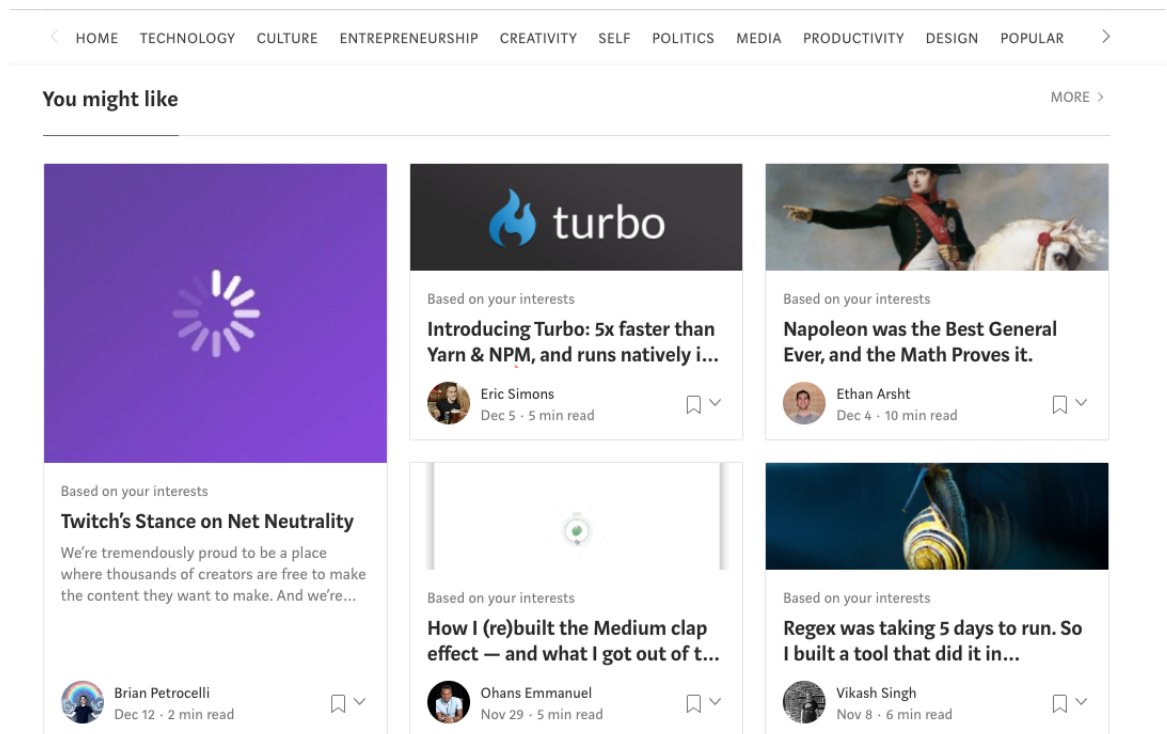


Figure 2: Medium recommended reads

4

Question

Answer

5

Question

11.11. Look at a sample of images or videos that have been tagged by users and separate the tags into three groups: those you think could eventually be done automatically by image processing and object recognition, those you think would not be possible to derive by image processing, and spam. Also decide which of the tags should be most useful for queries related to those images. Summarize your findings.

Answer

To answer this question I decided to use Instagram to find images tagged by users. Instagram, much like Twitter, allows for users to tag photos and posts using a # symbol. Its apparent that many posts on these sites bring in spam tags to get higher views or for other reasons.

For this problem I first started with a base hashtag “#lamborghini” and got a lot of unexpected results as shown in Figure 3. Just from the top six results, only half of the results were lamborghini model cars. You can tell just from these results that the non lamborghini tags were probably just spam tags.

When choosing one of the images from result page, such as Figure 4, its seen that the result is actually lamborghini was a spam tag for this image. The tags used were:

- Ferrari
- PaganiHuaryaBC
- pagani
- supercars
- fast
- speed
- london
- omg
- khk

- harrods
- city
- nightlife
- lamborghini
- astonmartin
- bugatti
- like4like
- likeforfollow
- followforfollow
- follow4follow
- rich
- wealth
- earth
- lifestyle

A majority of these tags would end up in spam. There are, however, a few of the tags that could be derived from image processing such as: supercars, bugatti, and city. If a image processing is done, assuming it is accurate, then the tags that address other types of cars, such as Ferrari, PaganiHuaryaBC, lamborghini, astonmartin, and paganica, should not be grouped together with the actual descriptors of this image. Tags that could not be taken from image processing, or rather should not, are things like wealth and earth since the first one is opinionated and earth is too broad of a topic. Tags such as likeforlike, follow4follow, omg, speed, wealth, and a few others, are more of spam tags for this image. The lamborghini tag itself is also considered spam for Figure 4, as the car isn't even a lamborghini.

The tags that would be most useful for queries would be the objects actually located inside the image. For example, bugatti is a good descriptor of the car in the image as its exactly whats in the image and it is also considered a supercar. However, if we search for supercar then we'll get much broader search results. So if I had to choose tags from this image, I

would probably only choose bugatti if I wanted concise results. Of course there are many tags that could be used for recommendations based off of this tag, such as rich, city, or supercar that could be recommended to users.

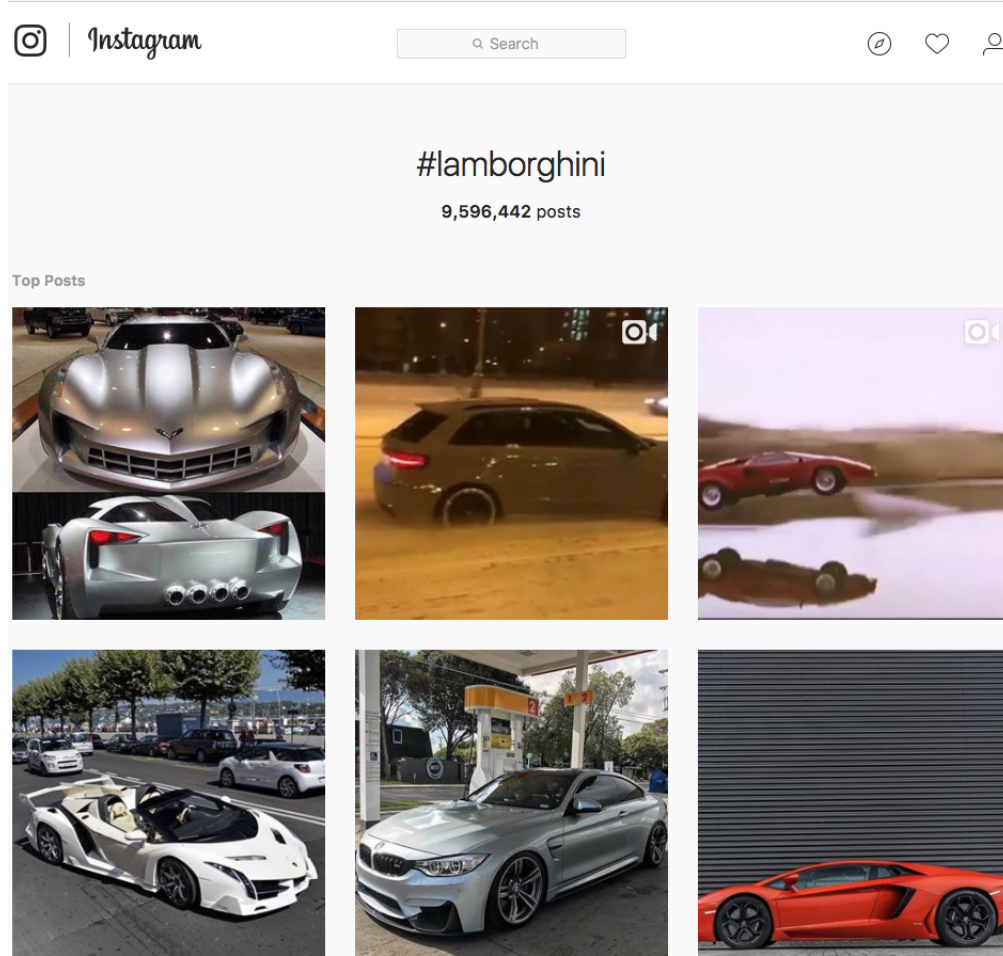


Figure 3: Results page from Instagram

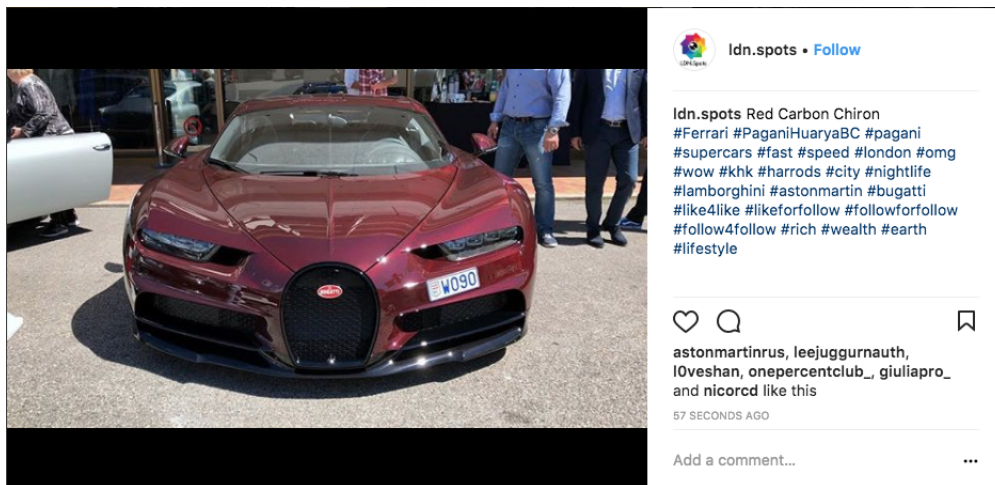


Figure 4: Selected result from Instagram for #lamborghini tag

References

- [1] Atkins, Grant. “CS734 Assignment 5 Repository” Github. N.p., 15 December 2017. Web. 15 December 2017.<https://github.com/grantat/cs834-f17/tree/master/assignments/A5>.