

CSC2231 – Internet Systems and Services

Paper Review – The Long Tail

Name: Alex Wun

Date: Nov. 27th, 05

This article discusses the implications of digital media distribution channels with respect to niche content. Traditional channels (physical products and retailers) incur manufacturing and distribution overhead. Additionally, consumers that would otherwise be interested in purchasing a certain product may be too far away from (or unable to find) a retailer or distribution outlet. Because of this, companies have traditionally focused on producing and selling the most popular content to ensure profitability. However, there are little to no (per product) production and distribution overheads in digital distribution. Combined with the fact that the physical location of customers is no longer a factor, it is now feasible and profitable to offer less-popular content from niche markets alongside the most popular content.

This is not a surprising result since niche contents have always existed and have always been profitable to some degree. The main barrier to entering niche markets has always been finding the customers that do exist, but may be geographically dispersed. The Internet masks this physical barrier.

There is also a secondary message in the article. By making niche content accessible, it is possible that some content that was previously “less-popular” becomes one of the most popular through increased exposure. In other words, the quality of search engines, content indexes and so on affects the popularity distribution. From this, I argue that the “sharpness” of a long-tailed popularity distributed is a reflection of the quality of the underlying searching/content discovery scheme. In other words, if some content provider had a terrible search scheme and its customers are only able to find the same handful of content, the popularity distribution would be sharply biased towards high-ranking objects. Few people will ever find the remaining content, causing the poorly indexed content to appear as statistically unpopular. But this is an artifact of their lousy content searching scheme. On the other hand, if the provider had a perfect content discovery scheme that could somehow read customer’s minds and return exactly (and only) the content they wanted, the popularity distribution would exhibit a fatter tail. i.e. There will be less of a contrast between the most popular and less popular content.

Web sites are organized as a hierarchy of pages: the home page links to several subsections which in turn have different sections and so on. If I am interested in a specific “leaf” web object, I will have to traverse the hierarchy of web objects starting at the root (the home page) to get to the content I want. Inevitably, every user will contribute a hit to the root content. But the root content isn’t popular per se! It only gets hits because it is the initial access point. A perfect scheme would allow me to access and download only the “leaf” web object I’m interested in. I believe that the popularity distribution would be a good measure of the quality of a search engine that indexes diverse content. I suspect (using only these hand-waving arguments) that a popularity distribution of web objects returned by a good search engine would be zipf with a low alpha (ignoring the search page itself). It would be interesting to see if this was the case for objects returned only through Google queries.