

# Correlation of Music Charts and Search Engine Rankings

Martin Klein  
Department of Computer  
Science  
Old Dominion University  
Norfolk, VA, 23529  
mklein@cs.odu.edu

Olena Hunsicker  
Department of Computer  
Science  
Old Dominion University  
Norfolk, VA, 23529  
koyal\_olena@yahoo.com

Michael L. Nelson  
Department of Computer  
Science  
Old Dominion University  
Norfolk, VA, 23529  
mln@cs.odu.edu

## ABSTRACT

We investigate the question whether expert rankings of real-world entities correlate with search engine (SE) rankings of corresponding web resources. We compare Billboards “Hot 100 Airplay” music charts with SE rankings of associated web resources. Out of nine comparisons we found two strong, two moderate, two weak and one negative correlation. The remaining two comparisons were inconclusive.

## Categories and Subject Descriptors

H.3.0 [Information Storage and Retrieval]:

## General Terms

Measurement, Performance, Design

## 1. INTRODUCTION

Billboard publishes the “Hot 100 Airplay”<sup>1</sup> music charts in the US each week based on radio airplay. The data is collected from more than 1000 radio stations and electronically monitored 24/7. Hence music charts (Billboard is considered the most authoritative) in some sense represent the notion of quality. Internet search engines (SEs) as well have become a powerful tool in our society. Top search results enjoy frequent hits whereas pages returned beyond the top 10 are rarely visited. Amento et al. [1] showed that web document quality can be estimated with hyperlink based metrics. But does that also mean, the quality of real world objects reflects in SE rankings? Given the power of charts and SEs we investigate the correlation between the two rankings.

## 2. THE RANKING COMPARISON

We took the Billboard charts from November 1st 2008 and extracted the names of the artists and bands. Often it is impossible to distill the most representative URL for a real world entity. We therefore map up to eight URLs to one artist or band listed in the charts. In order to compare this ranking with SE results we need to ensure the query to be unbiased. The queries consist of the URLs mapped to the artists concatenated with the *site:* (supported by Google) or *url:* (Yahoo and Live) query modifier. We queried the

<sup>1</sup>[http://en.wikipedia.org/wiki/Billboard\\_Hot\\_100\\_Airplay](http://en.wikipedia.org/wiki/Billboard_Hot_100_Airplay)

three SEs for the top 10, 25 and 50 results for a total of nine comparisons. Certain limits are enforced using SE APIs. Hence we must issue a series of overlapping queries to create an ordinal ranking of URLs relative to a specific SE. We used a variation of the strand sort algorithm and refer to [3] for details. A related experiment on the correlation of rankings of US college football teams can be found in [2]. We compute Kendall's Tau ( $\tau$ ) to determine the correlation between the charts and the ordinal ranking obtained from the SEs.

## 3. RESULTS

Table 1 shows the comparison results distinguished by SE. Each line shows  $\tau$ , the  $p$ -value and  $n$  indicating how many URLs associated with the artists were used to obtain this result. The statistically significant results are highlighted. Results provided by Google are all significant and show

	Top 10			Top 25			Top 50		
	$\tau$	$p$	$n$	$\tau$	$p$	$n$	$\tau$	$p$	$n$
Google	<b>0.82</b>	0.00	4	<b>0.71</b>	0.00	3	<b>0.53</b>	0.00	7
Yahoo	0.11	0.72	7	<b>0.32</b>	0.02	2	0.08	0.47	2
Live	<b>0.56</b>	0.03	5	<b>0.24</b>	0.09	3	<b>-0.21</b>	0.04	5

Table 1: Music Charts and SE Rank Correlation

strong correlations for the top 10 and 25 and a moderate correlation for the top 50 results. Yahoo and Live together provide one moderate and two weak correlations. Live shows the only negative correlation which underlines our intuition of an “inertia” implied in the web i.e., the ranking of real world objects changes too fast for the web to adapt its ranking. The high correlations shown by Google indicate that their ranking is influenced by a factor that captures the public focus faster than links, such as user queries en masse (cf. Google Zeitgeist<sup>2</sup>).

## 4. REFERENCES

- [1] B. Amento, L. Terveen, and W. Hill. Does “Authority” Mean Quality? Predicting Expert Quality Ratings of Web Documents. In *Proceedings of SIGIR '00*, pages 296–303, 2000.
- [2] M. Klein, O. Hunsicker, and M. L. Nelson. Comparing the Performance of US College Football Teams in the Web and on the Field. In *Proceedings of Hypertext '09*, 2009.
- [3] M. L. Nelson, M. Klein, and M. Magudamudi. Correlation of expert and search engine rankings. *arXiv*, abs/0809.2851, 2008.

<sup>2</sup><http://www.google.com/intl/en/press/zeitgeist2008/>