Comparing the Performance of US College Football Teams in the Web and on the Field

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
koval_olena@yahoo.com

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

ABSTRACT

In previous research it has been shown that link-based web page metrics can be used to predict experts' assessment of quality. We are interested in a related question: do expert rankings of real-world entities correlate with search engine (SE) rankings of corresponding web resources? To answer this question we compared rankings of college football teams in the US with rankings of their associated web resources. We looked at the weekly polls released by the Associated Press (AP) and USA Today Coaches Poll. Both rank the top 25 teams according to the aggregated expertise of sports writers and college football coaches. For the entire 2008 season (8/2008 – 1/2009), we compared the ranking of teams (top 10 and top 25) according to the polls with the rankings of one to eight URLs associated with each team in Google, Live Search and Yahoo. We found moderate to high correlations between the final rankings of 2007 and the SE ranking in mid 2008 but the correlation between the polls and the SEs steadily decreased as the season went on. We believe this is because the rankings in the web graph (as reported via SEs) have "inertia" and do not rapidly fluctuate as do the teams' on the field fortunes.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: [Information Search and Retrieval]

General Terms

Measurement, Performance, Design

Keywords

Search Engines, Ranking, Correlation, Real World Objects

1. INTRODUCTION

As a society, we enjoy lists, presumably compiled by "experts", that rank items, events, people, places, etc. At best, these lists are informative and help convey notions of quality in a compact manner. At worst, these lists can be misleading, biased, or overly simplified. Regardless, lists proclaiming the top 10, 25 or 50 of various resources are a persistent part of our culture.

At the same time, search engines now play a central role in society. The "big 3" search engines (SEs) – Google, Live (formerly MSN), and Yahoo – are the primary tool for discovering web resources for many people. Acquiring a high ranking in SEs is so important that an entire discipline and economy of search engine optimizers (SEOs) has developed to help people raise the ranking of their web pages. Thus SEs move from a simple navigation and discovery aid to a powerful cultural force. In some sense, if a web page does not appear in the first few pages of a SE's result set for a particular query, it is as if it does not exist at all.

Given the power that expert lists and SEs have, we are interested in their intersection. In particular, we want to know if expert rankings of "real-world" resources such as collegiate football programs in the United States that change on a weekly basis during the season correlate to the search engine rankings of their corresponding web resources. It was our intuition that highly ranked real-world resources would be correspondingly highly-ranked in SEs.

To answer this question we used the Associated Press (AP) Poll and USA Today Coaches Poll which publish each week the top 25 teams according to their constituents. Since it is hard to argue for one canonical URL per football program we mapped each team to up to eight URLs ($n = 8$) and created an ordinal ranking of the URLs in a SE independent of any keyword query. We first investigated the correlation between the final rankings from the previous season (2007) as well as the pre-season rankings (from August 2008) and SE ranking taken in August 2008. We further investigated the correlation of the rankings published once a week during the season as well as final season rankings with the SE rankings from the according weeks. The college football season started in the last week of August 2008 and lasted 15 weeks. We used Kendall's Tau (τ) to test for statistically significant ($p < 0.05$), moderate ($0.40 < \tau \leq 0.60$) or strong ($0.60 < \tau \leq 0.80$) correlations between the expert rankings and SE rankings.

2. RELATED WORK

There have been a number of studies that assess the quality of SE results for a particular query, but relatively few that examine the ranking of URLs with respect to the rank

HT'09, June 29–July 1, 2009, Torino, Italy.

or status of their corresponding "real world object" (e.g., a company, person or university).

2.1 Quality and Authority in the Web

"Does 'Authority' mean Quality?" is the question Amento et al. [1] asked when they evaluated the potential of link- and content-based algorithms to identify high quality web pages. Human experts rated web documents from the Yahoo directory related to five popular topics by their quality. Amento et al. found a high correlation between the rankings of the human experts leading to the conclusion that there is a common notion of quality. By computing link-based metrics as well as analyzing the link neighborhood of the web pages from their dataset they were able to evaluate the performance of machine ranking methods. Here too they found a high correlation between in-degree, Kleinberg's authority score [9] and PageRank. They isolated the documents that the human experts rated with good quality and evaluated the performance of algorithms on that list in terms of precision at 5 and at 10. In-degree e.g., has a precision at 5 of 0.76 which means on average almost 4 of the first 5 documents it returns would be rated good by the experts. In general they find that in-degree, authority score and PageRank are all highly correlated with rankings provided by experts. Thus, web document quality can be estimated with hyperlink based metrics.

Bharat and Mihaila [4] propose a ranking scheme based on authority where the most authoritative pages get the highest ranking. Their algorithm is based on a special set of "expert documents" which are defined as web pages about a certain topic with many links to non-affiliated web pages on that topic. Non-affiliated pages are pages from different domains and with sufficiently different IP address. These expert documents are not chosen manually but automatically picked as long as they meet certain requirements (sufficient out-degree, etc). In response to a user query the most relevant expert documents are isolated. The proposed scheme locates relevant links within the expert documents and follows them to identify target pages. These pages are finally ranked according to the number and relevance of expert documents pointing to them and presented to the end user. Bharat and Mihaila evaluated their algorithm against three commercial search engines and found that it performs either just as good or in some cases even better than the top search engine when it comes to locating the home page of a specific topic. The same is true for discovering relevant pages to topic (where many good pages exist).

Rieh [14] conducted a study on user's judgment of information quality and cognitive authority in the web by observing the user's searching behavior. The idea was to understand the factors that influence user's judgment of quality and authority in the web. In her work information quality on an operational level is defined as "the extend to which users think that the information is useful, good, current and accurate". Cognitive authority is "the extend to which users think that they can trust the information". Rieh found that users do predictive judgment (before opening the page) and evaluative judgment (after opening the page) when it comes to the choice what page and item on a page to look at. If the evaluative judgment does not correlate with the expectations made in the predictive judgment the user usually starts a new page or goes back to a previous one. If the two judgments match however the user stays on the page and uses its information. She also found in her experiments that users identify the facets characterizing cognitive authority in the web as: trustworthiness, reliability, scholarliness, credibility, officialness and authoritativeness. However for the subjects she conducted the study with authority was more important for some search tasks than for others. Looking for medicine e.g., authority was a major concern but did not affect the subjects much for the travel research task.

Rieh and Belkin [15] conducted a similar study about people's decision making in respect to information quality and cognitive authority in the WWW. This study confirms the intuition that users of the web assess information quality based on source credibility and authority. Authority can be seen on a institutional level e.g., academic or governmental institutions and on a personal level e.g., professional experts. Another interesting finding of this work is that users believe that the web is less authoritative and also less credible than other, more conventional information systems.

Capra et al. [5] found that during the campaign preceding the presidential election in the US is 2008 Internet resources such as YouTube videos and blogs retrieved as the result of various search queries were highly relevant not only to the general topic but also to the candidates (real world objects) themselves. They further show that topic relevant resources (YouTube videos) can be obtained from secondary sources such as blog entries.

2.2 Quality as a Factor in Web Page Ranking

Cho et al. [7] observe a "rich-get-richer" phenomenon where popular pages tend to get even more popular since search engines repeatedly return popular pages first. As other studies by Cho [6, 13] and Baeza-Yates [3] have shown, PageRank is significantly biased against new (and thus unpopular) pages which makes it problematic for these pages to draw the user's attention even if they are potentially of high quality. That means the popularity of a page can be much lower than its actual quality. Cho et al. propose page quality as an alternative ranking method. By defining quality of a web page as the probability that a user likes the page when seeing it for the first time the authors claim to be able to alleviate the drawbacks of PageRank. With the intuition from PageRank that a user that likes the page will link to it the algorithm is able to identify new and high quality pages much faster than PageRank and thus shorten the time it takes for them to get noticed.

2.3 Quality of Web Documents

Lim et at. [10] introduce two models to measure the quality of articles from an online community like Wikipedia without interpreting their content. In the basic model quality is derived from the authority of the contributors of the article and the contributions from each of them (in number of words). The peer review model extends the basic model by a review aspect of the article's content. It gives higher quality to words that "survive" reviews.

An approach to automatically predict information quality is given by Tang et al. [16]. Analyzing news documents they observe an association between users quality score and the occurrence and prevalence of certain textual features like readability and grammar.

2.4 Ranking of URLs of Real World Objects

Upstill, Craswell and Hawking [17] studied the PageRank and indegree of URLs for Fortune 500 and Fortune Most Admired companies. They found companies on those lists averaged 1 point more PageRank (via the Google toolbar's self-reported 0-10 scale) than companies not on the list. They also found that IT companies typically had higher PageRank than non-IT companies. Similar to [1], they found indegree highly correlated with PageRank.

In previous work we found few correlations between expert rankings of "real world objects" (e.g., top schools according to US News and World Report, Fortune 500 companies, tennis players ranked by the ATP) and the ranking of corresponding URLs in SEs [12]. Limitations of the previous study included that we only accounted for a single URL per real world object (i.e., n=1) and the real world objects that were studied were not necessarily well represented in the popular culture. In the current study we chose college football because of its popularity and is well-defined start and end points.

As part of our later, more complex study [8] however we found (statistically significant) strong and moderate correlations between US music charts and the SE rankings of corresponding web pages of the artists and bands. We used Billboard's "Hot 100 Airplay" music charts and map up to eight URLs to the corresponding artists and bands.

3. EXPERIMENT DESIGN

The following section details the chosen expert lists, explains how we chose URLs to correspond with the entries in the expert lists, and discusses the searching and ranking algorithms and other operational details.

3.1 Choosing Expert Lists

For this experiment we chose the Associated Press (AP) and USA Today Coaches Polls, which are the two most popular polls (i.e., expert lists) for the National Collegiate Athletic Association (NCAA) Football Bowl Subdivision (FBS; formerly known as "Division I-A"). There were 119 teams competing in the FBS during the 2008 season. The AP Poll consists of 65 sportswriters and broadcasters that vote weekly with their top-rated choice receiving 25 points and the 25th rated choice receiving 1 point. The USA Today Coaches Poll is scored similarly and has 63 college football head coaches participating. We also examined the less popular Harris Interactive Poll and Massey Rankings (which are functionally similar to the polls) but their data is incomplete with the Harris Interactive Poll not beginning until week four of the season and the Massey Rankings not being issued for the final week of the season. Their results were similar to those of the AP and USA Today and as such are not reported here.

The accuracy, criteria or bias of these rankings may be critiqued, but that is not the purpose of this investigation. We simply accept the rankings as given from the experts in the polls. A full discussion of college football and how the rankings are used to determine championship eligibility is beyond the scope of this paper.[1]

The 2008 season began on August 28, 2008 and concluded on January 8, 2009. We collected 18 instances of poll data: the final polls from the 2007 season (as a baseline), the 2008 pre-season polls, and then once for each of the 16 weeks of the season.

[1]Please see `http://en.wikipedia.org/wiki/College_football` for a history and explanation of college football and `http://en.wikipedia.org/wiki/Bowl_Championship_Series` for the poll results, explanation of polls, championships and associated controversy.

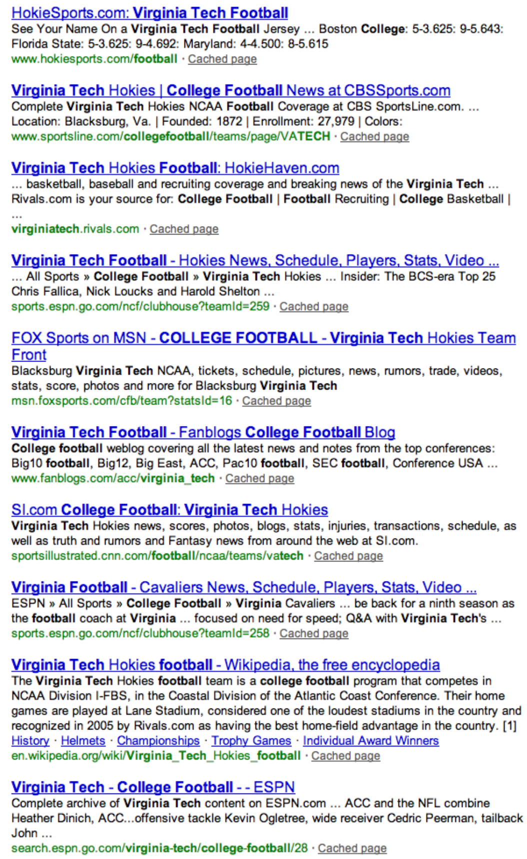


Figure 1: Top 10 Search Results for "Virginia Tech College Football" in Live

3.2 Mapping Resources to URLs

After the expert lists have been chosen, we began the process of mapping their real-world objects to URLs. It is not trivial, if not impossible, to assign one URL to a school's football program. There is probably only one official website hosted by the college itself but what about fan sites or commercial sports sites like ESPN? This is a problem for all hyperlink derived metrics: multiple candidate URLs can compete for a limited number of links on web pages, thereby reducing their importance or popularity metric. Figure 1 shows a screen shot of the result page when querying `Live.com` for *Virginia Tech College Football*. Note that the Virginia Tech athletes are called *Hokies*. We can see a variety of pages in the result set such as the official site `http://www.hokiesports.com/`, commercial sports sites such as `http://sports.espn.go.com/ncf/clubhouse?teamId=259`, the Wikipedia page `http://en.wikipedia.org/wiki/Virginia_Tech_Hokies_football` and even fan blog sites. Since we do not know which of these sites an author will link to when creating a link to the concept of "Virginia Tech Football" and we can not dictate which site is the most representative for the concept we have to treat them (at least the top n) in aggregate.

Since our experiment covers the entire football season of 2008, we need to consider the temporal aspect in SE queries and their resulting ranks. The temporal shift in SE queries was recently shown by Backstrom et al. [2] and can annually be followed by observing the Google Zeitgeist data[2]. We re-queried the three major SEs on a weekly basis throughout the season in order to obtain the mapping URLs per school. The query consists of general terms following the pattern *schoolname+College+Football* for example *Ohio+State+College+Football*. We sample the top eight URLs from the result set, dismissing URLs that contained unescaped white spaces, unescaped unsafe characters and URLs with more than one parameter such as

`http://www.foo.bar/?parameter1=a&parameter2=b.`

The reason for that is they are ignored by SE APIs when querying them for indexed or cached URLs. Table 1 shows the top three URLs mapped to the 12 colleges that appear in all AP polls for the entire season. The attendance data was obtained from the NCAA website[3].

3.3 Creating an Ordinal Ranking of URLs from SE Queries

We developed a Perl program that takes a list of URLs and queries search engines to determine their relative ordering of those URLs. We do not determine a search engine's absolute ranking for any particular URL. That is, we do not compute:

$rank(URL_A) = 0.92$
$rank(URL_B) = 0.73$
$rank(URL_C) = 0.42$
...

We also are not interested in estimating the PageRank (or related metrics), independent of SEs, through link neighborhoods or other means: the SEs are the subject of our study, not the web graph itself. Instead, using a variation of strand sort (illustrated in section 3.3.2), we simply determine that a search engine ranks the URLs in order:

$$rank(URL_A) \geq rank(URL_B) \geq rank(URL_C) \geq ...$$

Note that the ranks of both the experts and search engines are ordinal variables, so generally:

$$distance(URL_A, URL_B) \neq distance(URL_B, URL_C).$$

The program queried the APIs of Google, Live and Yahoo. Although it has been shown that search engine APIs return different results than the public (human) interfaces [11] and possibly use a smaller index, we chose to use the APIs instead of "page-scraping" the results to avoid being denied access by the search engines.

Although the SE APIs can be queried for backlinks or ranking metrics, previous research has shown that these values are not always accurate, perhaps intentionally so to prevent reverse engineering of SE ranking algorithms [11]. Note that it is not our goal to compute the interval value of a particular URL in a given SE, but rather just to produce an ordinal ranking of URLs for a SE. We treat the SEs as a black box ranking system and do not try to reverse engineer its hyperlink-based methods.

We need to point out that all the expert lists and the SE APIs queried are biased toward the English language and since the ranked items (football programs) are primarily of interest to the United States, we made no attempt to query non-English language SEs.

Ideally, we could submit all URLs to a SE in a single query and record the resulting ordering. However, each SE has query length limitations for both characters and terms and queries that exceed these limitations are silently truncated. Google for example allows only 1000 queries per day and the query length must not exceed 2048 bytes and 10 words. We must issue a series of overlapping queries to create an ordinal ranking of URLs relative to a specific SE. To this end, we used a variation of strand sort[4]. Strand sort is a sorting algorithm that uses multiple intermediate data structures to temporarily store a sorted subset of the data. These structures are eventually gathered together to sort the entire list of data. This behavior makes it part of the family of distribution algorithms.

3.3.1 Querying Search Engine APIs

In order to determine the SE ranking of the URLs we must form unbiased queries. We do that by using the `site:` query modifier which is supported by Google and the `url:` modifier, supported by Yahoo and Live. It works as a filter by restricting the results to websites in the given domain only. We query for several URLs simultaneously (specified by q) and thus combine the URLs and the `site:` or `url:` modifier with the boolean `OR` operator (which is supported by all three search engines). This boolean operator returns results that match either side of the query string divided by the `OR`. Since our queries consist of URLs only, each with the same modifier and combined with the boolean operator and no keywords added, all search results have theoretically an equal opportunity to be returned as the top result and "only" the search engine's ranking is dictating the ranking of the URLs now. We verified that the search results were commutative: the order of the URLs in the queries did not change the final rankings. As an example, the query for the first five programs from Table 1 using the first URL per school only ($n = 1$) would be:

```
site:http://usctrojans.cstv.com/sports/m-footbl/usc-m-
footbl-body.html OR
site:http://uga.rivals.com/ OR
site:http://sportsillustrated.cnn.com/football/ncaa/
teams/ohiost/ OR
site:http://www.soonersports.com/ OR
site:http://www.gatorzone.com/
```

3.3.2 An Example Ordinal Ranking of URLs

We illustrate creating an ordinal ranking of URLs with an example. Assume an unsorted list UL with eight URLs (G, E, B, A, C, H, F, D). The expected outcome in the sorted list SL will be ranked in lexicographical order and we chose $q = 3$. The first q URLs (G, E, B) are queried against the search engine an the result is sorted (B, E, G). The overlap URL (the q^{th} element), let us call it OL, is the URL G since

[2] `http://www.google.com/intl/en/press/zeitgeist2008/united_states.html`
[3] `http://web1.ncaa.org/d1mfb/Internet/attendance/IA_AVGATTENDANCE.pdf`
[4] `http://en.wikipedia.org/wiki/Strand_sort`

Rank	School	URL	Avg Attendance (Rank)	Attendance Capacity (%)
1	USC	http://usctrojans.cstv.com/sports/m-footbl/usc-m-footbl-body.html http://usctrojans.cstv.com/ http://deadspin.com/5042455/college-football-previews-2-usc	87476 (9)	95.08
2	Georgia	http://uga.rivals.com/ http://sportsillustrated.cnn.com/football/ncaa/teams/georgi/ http://sportsillustrated.cnn.com/football/college/teams/ggb/	92746 (5)	100.00
3	Ohio State	http://sportsillustrated.cnn.com/football/ncaa/teams/ohiost/ http://www.usatoday.com/sports/college/football/bigten/osu.htm http://msn.foxsports.com/cfb/team?statsId=33	105110 (3)	102.72
4	Oklahoma	http://www.soonersports.com/ http://www.sportsline.com/collegefootball/teams/page/OK http://sportsillustrated.cnn.com/football/ncaa/teams/okla/	84858 (11)	103.34
5	Florida	http://www.gatorzone.com/ http://www.gatorzone.com/football/ http://www.sportsline.com/collegefootball	90388 (8)	102.08
6	Missouri	http://www.sportsline.com/collegefootball/teams/page/MO http://sportsillustrated.cnn.com/football/ncaa/teams/missou/ http://mutigers.cstv.com/	60232 (31)	88.12
7	Texas	http://www.texascollege.edu/football.htm http://www.sportsline.com/collegefootball/teams/page/TX http://www.mackbrown-texasfootball.com/	85144 (10)	100.02
8	Texas Tech	http://www.sportsline.com/collegefootball/teams/page/TXTECH http://sportsillustrated.cnn.com/football/ncaa/teams/txtech/ http://deadspin.com/5038247/college-football-previews-14-texas-tech	51911 (40)	98.16
9	Alabama	http://www.rolltide.com/ http://www.sportsline.com/collegefootball/teams/page/AL http://sportsillustrated.cnn.com/football/ncaa/teams/alabam/	92138 (7)	100.00
10	Brigham Young	http://www.sportsline.com/collegefootball/teams/page/BYU http://sportsillustrated.cnn.com/football/ncaa/teams/byu/ http://www.usatoday.com/sports/college/football/mwest/byu.htm	64497 (27)	100.71
11	Penn State	http://www.usatoday.com/sports/college/football/bigten/psu.htm http://sportsillustrated.cnn.com/football/ncaa/teams/psu/ http://thequad.blogs.nytimes.com/2008/11/08/penn-state-loses-but-college-football-may-win/	108917 (2)	101.52
12	Utah	http://utahutes.cstv.com/sports/m-footbl/utah-m-footbl-body.html http://utahutes.cstv.com/ http://sportsillustrated.cnn.com/football/ncaa/teams/utah/	42593 (54)	93.34

Table 1: The Top Three Mapped URLs for the 12 Schools that Overlap in all AP Pools Throughout the Entire Season

it is the result with the lowest rank in this subset of URLs. The other two URLs (B, E) are stored in SL.

In the next iteration we pull the next $q-1$ elements from UL and together with $OL = (G)$ form a new query (G, A, C) for the search engine. The result is (A, C, G) indicating that A and C can be ranked anywhere higher than OL and thus need to be merged with the elements in SL. First we take A and query it together with (B, E) and get the result (A, B, E). Since SL contains just these three elements we are assured we found the correct rank for A. We know that C was ranked lower than A and thus only need to query C together with all elements from SL ranked below A. Thus we query (C, B, E) and receive the result (B, C, E) which we can append to the top ranked result A. SL now consists of (A, B, C, E). G remains the OL since it was still the lowest ranked element in the subset and will now (in the third iteration) be queried together with the next $q-1$ elements from UL. The query (G, H, F) returns (F, G, H) which means H as the lowest ranked URLs will become the new OL and F and G need to be merged with all elements of SL. First we query F together with the first $q-1$ elements from SL and get the result (A, B, F). This may not be the final position of F yet since SL contains more than three elements. All we know at this stage is that F is ranked below A and B. Thus we need to also query (F, C, E) and will get (C, E, F). Now all elements in SL are checked against F and it turns out F is the last element and thus can be appended to SL which now holds the ranking (A, B, C, E, F).

As the second part of this third iteration we need to find the final position of G. We again know its ranked lower than F and since F is the last element of SL we can simply append G to SL which now contains the sorted list (A, B, C, E, F, G). The new OL is queried together with the remaining element of UL, D and the query returns (D, H). This result tells us we need to treat D the same way like we did with F in the third iteration. We query (D, A, B) and get (A, B, D) then we query (D, C, E) and get the result (C, D, E). Now we have determined the final position of URL D and can place it accordingly in SL. Since the OL is still H and UL is empty we are assured H is the lowest ranked URL in the entire set and can simply append H to SL. This is the final step of the algorithm and SL now holds the sorted list containing all URLs (A, B, C, D, E, F, G).

More details on the algorithm can found in the technical report [12].

3.4 Weighting Ranked URLs

If we chose to map a real world resource, in our case a collegiate football program, to more than one URL ($n > 1$) we need to accumulate the ranking score for each and every single one of the URLs in order to compute one overall score for the school's program. We assign weights per URL depending on its rank using the following formula:

$$Weight \; = \; 1 - \frac{P}{T} \tag{1}$$

Where P is the position of the URL in the result set and T is

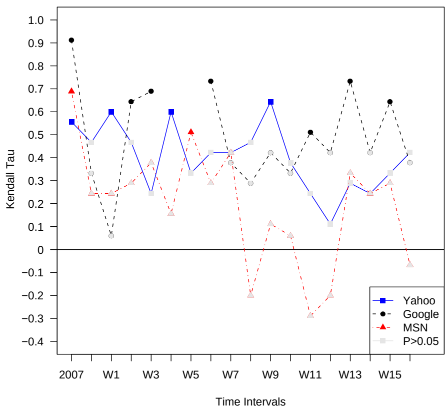


Figure 2: Rank Correlation of Top 10 Schools from the AP Poll and Google, Yahoo and Live

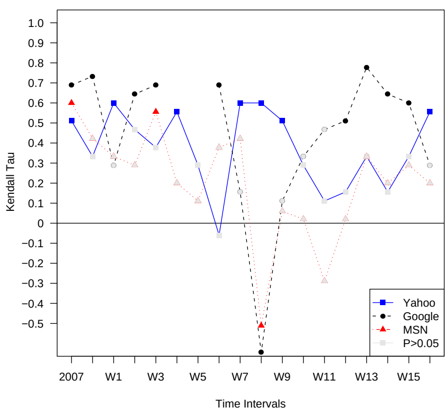


Figure 3: Rank Correlation of Top 10 Schools from the USA Today Poll and Google, Yahoo and Live

the total number of URLs in the list which is equal to n times the number of football teams. By adding the weights for all URLs mapped to one team, we can rank all teams by their accumulated weight. We therefore have applied a weighting scheme for multiple URLs per entity and can compute the correlation between entities from SE based and poll based rankings.

4. RESULTS

4.1 Correlations Between Polls and Search Engine Rankings

Figure 2 shows the correlations measured in Kendall τ between the top 10 ranked schools based on the AP Poll and SE rankings. Figure 3 shows the correlation between the USA Today poll and the SE rankings. Figures 4 and 5 show the data for the top 25 schools and the according polls. In all four figures progress in time is represented on the x-axis. Google data for weeks four and five is missing due to corrupted code. The error was noticed and fixed on

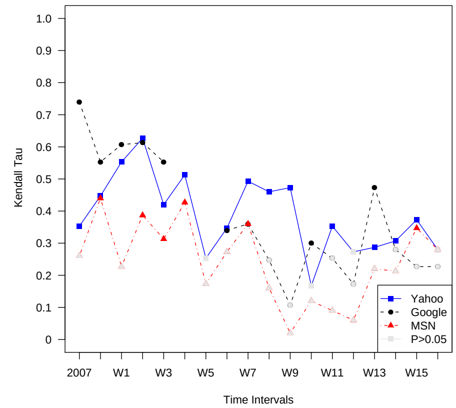


Figure 4: Rank Correlation of Top 25 Schools from the AP Poll and Google, Yahoo and Live

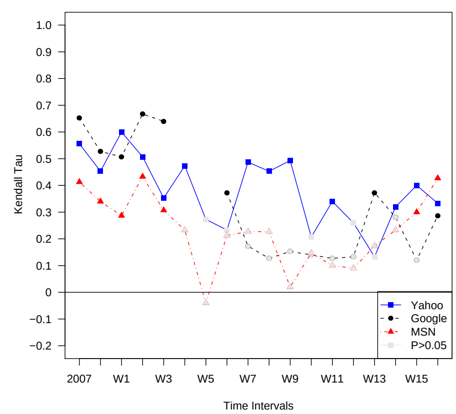
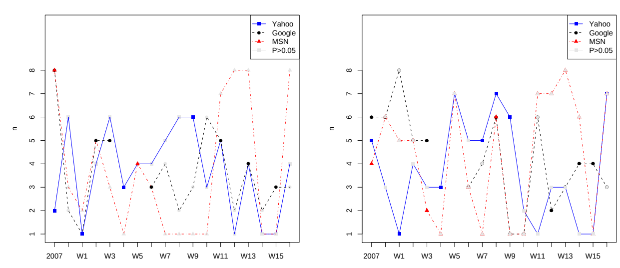
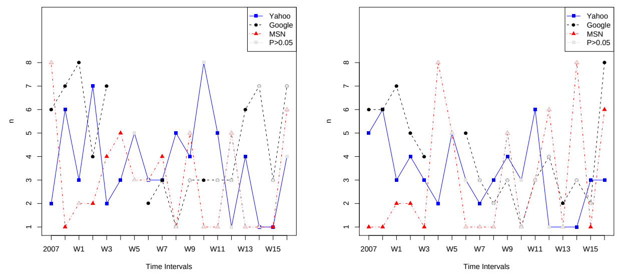


Figure 5: Rank Correlation of Top 25 Schools from the USA Today Poll and Google, Yahoo and Live



(a) AP Poll (b) USA Today Poll

Figure 6: n Values for Correlation of Top 10 Schools with Google, Yahoo and Live Ranking



(a) AP Poll (b) USA Today Poll

Figure 7: n Values for Correlation of Top 25 Schools with Google, Yahoo and Live Ranking

time to determine the data for week six. The first, leftmost data point in the graphs shows the correlation between the final rankings of the 2007 season and SE ranking taken in August 2008. All other data points show the correlation between polls and SE rankings obtained at the same time starting with the pre-season rankings in August of 2008 and ending with the final rankings of the 2008 season. The different line styles and colors represent the results from the three SEs and all highlighted dots (in the according color) indicate a statistically significant correlation with $p < 0.05$. The correlation for the top 10 results generally seems slightly higher. A high correlation can be found for the final rankings of the previous season. The Google ranks for example show a high correlation with the AP polls of 2007 with $\tau > 0.9$ in Figure 2 and $\tau = 0.75$ in Figure 4. The correlation however decreases for the post-season polls and drops further with the ongoing season. We explain this pattern with the implied "inertia" in the web. The poll based rankings change frequently and the web can not catch up or in other words, the real world moves faster than the web can adjust. Note that we do not distinguish between n-values in Figures 2 through 5. Since these graphs are geared towards showing the correlation, we chose to display the highest correlation values available, regardless of the n-value.

Figures 6 and 7 show the same data that was used to obtain the data shown in the Figures above but this time distinguished by the n-value. Here too the solid colored dots indicate statistically significant results. Especially for the top 25 schools n-values between two and six seem to produce the most significant results. The described pattern repeats for Harris and Massey polls hence we chose not to show the graphs here.

4.2 Evaluating Correlation of Overlapping URLs Over Time

We compare up to eight URLs mapped to each of the top 25 ranked football programs. However, only 12 of these programs occur in all polls throughout the entire season (see Table 1). The goal of this experiment is to see whether SE rankings trail poll based rankings by a certain amount of time. The intuition is that the web has an implied "inertia" and we therefore could see a somewhat delayed correlation for the 12 overlapping programs. We declared the AP poll based ranking of each week as a separate "truth value" and computed the correlation between eight URLs mapped to the 12 overlapping schools ranked per that particular truth and SE rankings. The result is 15 comparisons, one for each week since each week represents one truth value. If the SEs would trail the polls, we would expect to see an increased correlation in the weeks following the truth value. Figure 8 shows all 15 graphs where the solid black line represents the τ values and the dotted line the p-value. The big square blue dot on the solid line indicates the current truth value. From the graphs we can see a moderate correlation in the first few weeks but as the season continues and the performance of the teams diverges, the correlation decreases. This again can be explained with the dynamics in the polls. The on-field performance which is reflected in the poll based rankings fluctuates too much for the web to adjust its ranking.

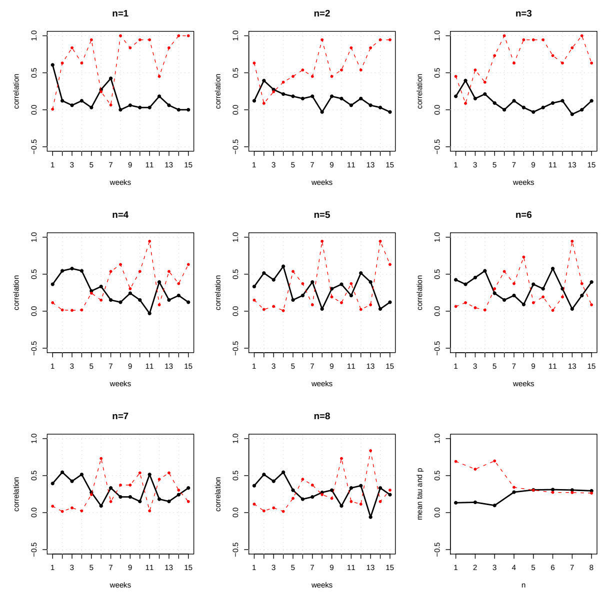


Figure 9: Correlation of AP Poll Data and Google Based Ranking with Varying n

4.3 Evaluating n-Values

We offer a maximum of eight representative URLs per entity. The number is convenient since the Google AJAX API offers a maximum of eight results per query. The question now is, does the correlation change with fluctuating n and if so, is there an n-value performing best? Figure 9 shows the correlation of Google rankings with AP poll data with n values varying between one and eight. Similar to the concept of truth from Figure 8 we assume here week one to be the truth and compute its correlation with SE rankings of the consecutive weeks. We again can see a moderate correlation in the early weeks and a drop in the weeks following regardless of the n value. The plot in the lower right hand corner displays the mean τ- and mean p-values for $1 \leq n \leq 8$. By visual observation it seems impossible to determine an optimal n-value. Figure 10 displays n-values for all correlations taking week one through 15 as the truth. In the first couple of weeks n-values greater or equal than four seem to perform best and if week three and four is considered the truth, an n-value of four or five seems best. From week five on however, a value of $n = 3$ shows the best mean performance. This result seems to imply that there indeed is a sweet spot

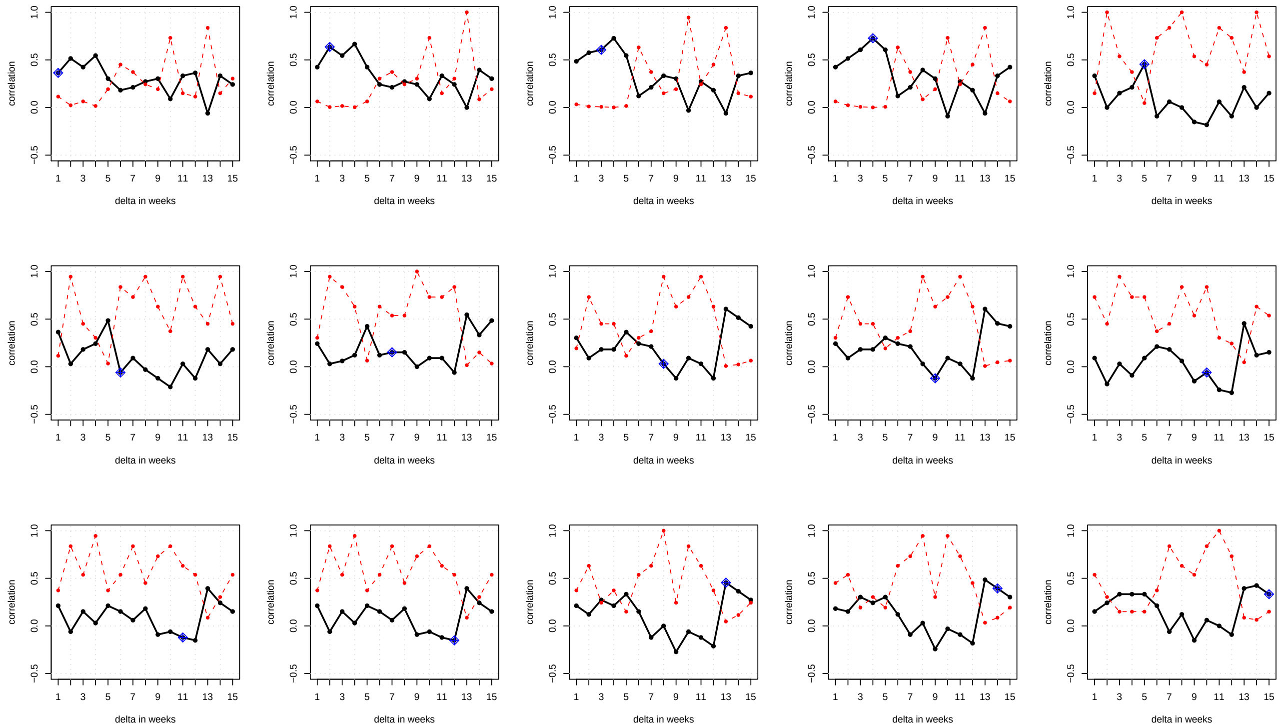


Figure 8: Correlation of Eight URLs Mapped to the Top 25 Schools from the AP Poll and Google Ranking with Different "Truth Values" Over Time

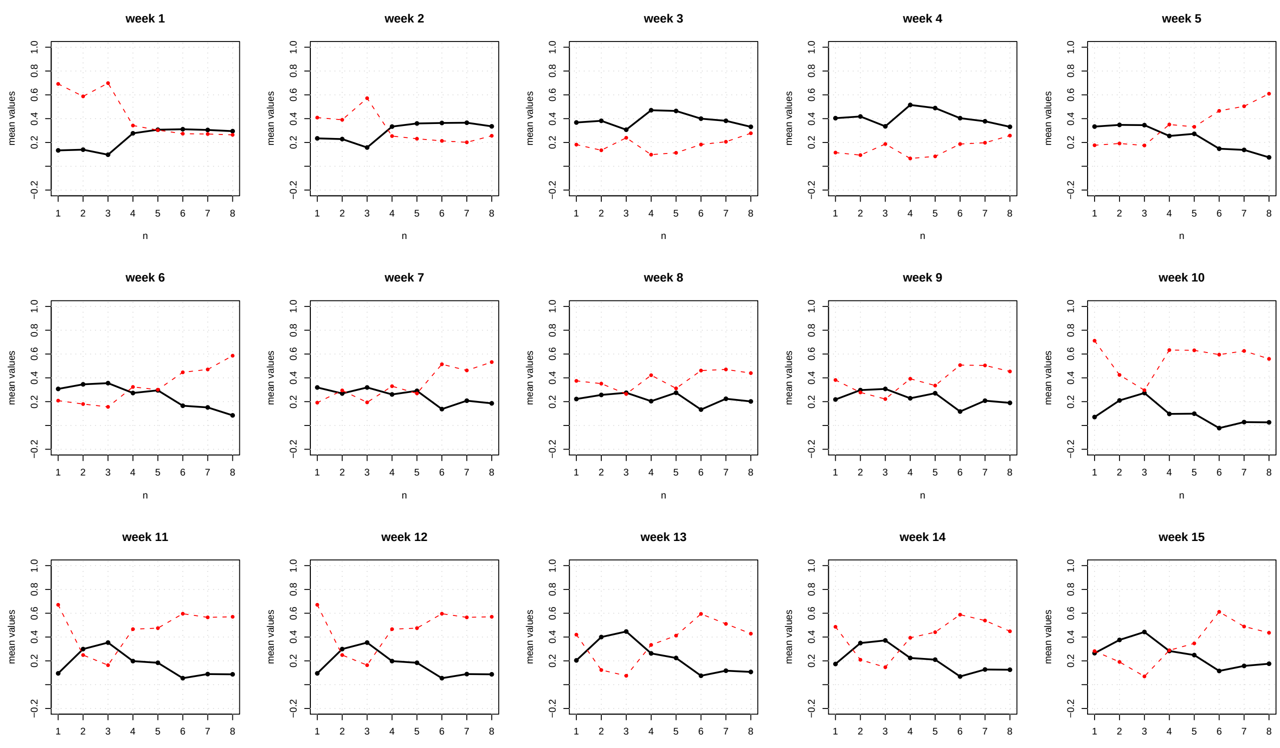


Figure 10: Comparison of Mean τ- and p-Values for Varying n

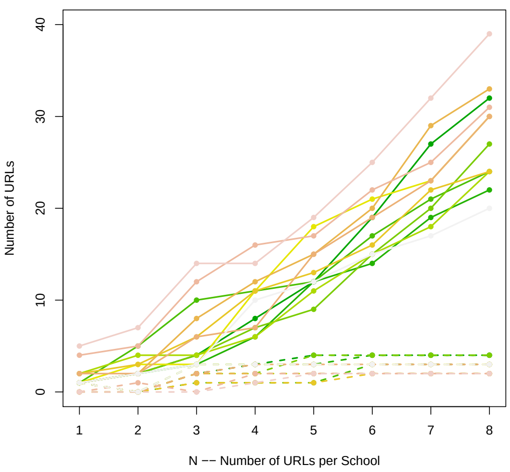


Figure 11: Union and Intersection of Mapping URLs Throughout the Season

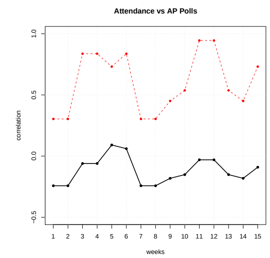
for n. Choosing $n < 3$ is not sufficient and $n > 5$ also seems to hurt rather than benefit the correlation. The aggregated score for the n representative URLs does not seem to improve drastically with $n > 5$ and is obviously too low for $n < 3$. Therefore choosing n between three and five seems preferable for a moderate correlation between SE rankings and frequently changing poll based rankings.

4.4 Change in Mapping URLs

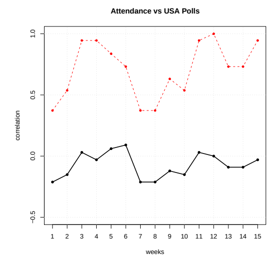
As mentioned above, we re-query SEs each week to determine which URLs are assigned to which school. An interesting question is, depending on n, how does the set of URLs change over the period of the season. The intuition is that, especially with a high n-value, we will see a constant core of URLs and a few URLs swapping in and out of the set. Each of the solid lines in Figure 11 displays the union of URLs mapped to one of the 12 overlapping schools. The dotted lines represent the intersection. It is not important to identify a particular school in the graph hence the lines are rather indistinguishable. The intersection of URLs does not go above four but interestingly does not increase anymore in our sample set for $n \geq 6$. So for sufficiently large n the core of URLs here seems to be of size four. For the union we see a steady increase with larger n-values. Intuition tells us that with n much greater than eight the lines would eventually level off meaning fewer and fewer new URLs would be acquired for the mapping.

4.5 Correlation Between Attendance and SE and Poll Ranking

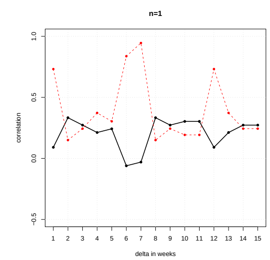
We also investigated the correlation of the school's ranking based on home game attendance and SE as well as poll based rankings. The ranks regarding the attendance is shown in Table 1. The intuition is that large schools with a large stadium and a broad alumni community enjoy high SE ranks regardless of their on-field performance. The University of Michigan for example leads the attendance ranking with an average of more than 110,000 but (at least in the 2008 season) did not have an impact in the polls. Figures 12(a)



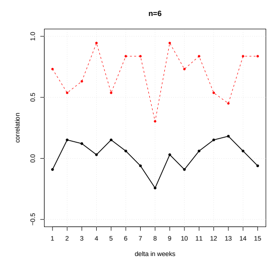
(a) AP



(b) USA Today



(c) Google, $n = 1$



(d) Google, $n = 6$

Figure 12: Correlation between Attendance Rankings and AP, USA Today Polls and Google SE Rankings

and 12(b) show the measured correlation between the attendance ranking and the AP and USA Today polls. As expected, there is no statistically significant correlation with the p-value constantly above 0.05. The correlation with SE rankings in dependence of n is shown in Figures 12(c) and 12(d). Due to space restrictions we chose to display graphs for $n = 1$ and $n = 6$ only as they represent the observed correlation for all n. We expected a higher correlation than we found. However, the size of a football stadium is not elastic meaning even if all games are completely sold out, the school may not make it into the higher ranks if the stadium is just not big enough. The score of the web ranking in contrast is elastic, so each school can theoretically make it to the top. Our intuition is that the attendance ranking may give an indicator for mid and lower level schools since everyone loves winners and therefore the games of the 12 mentioned schools are mostly sold out.

5. CONCLUSIONS AND FUTURE WORK

Inspired by the question of Amento et al. [1] "Does Authority mean Quality?", we have asked "Does Quality mean Authority?" In previous work we have been unable to verify that "real world" quality correlates with web-based "authority" [12]. In this study, we address this question by comparing two expert rankings, the AP and USA Today Top 25 Polls, of NCAA FBS football programs against SE rankings. These tests were conducted before and during the 2008 season. In our tests we see statistically significant high and moderate correlations for the last seasons final rankings and rankings early in the season. With the season progressing and the fortune of the teams changing however, the correlation decreases because of "inertia" in the web. In the off-season, URLs describing teams anticipated to perform well during the season have enough time to acquire the links necessary to drive up their SE rankings. But as the season unfolds over the course of 12-15 games, their acquired links are not modified rapidly enough to reflect their current standings.

Large public schools with a traditionally powerful football program seem to keep their high ranking in the web, regardless of their on-field performance. We initially expected to find a correlation between rank and attendance, but for the top programs attendance is inelastic. Attendance is governed by stadium size and is frequently at or near capacity. Attendance may be a better indicator for the bottom 25 teams than the top 25.

Mapping from real-world objects to corresponding web pages can be difficult and we therefore queried SEs for up to eight representative URLs per team and computed an aggregate weighting score to rank them accordingly. Although the URLs were the top returned results for each query, perhaps the queries were not the best. For example, in future work we could include program nicknames (e.g., "Hokies" for Virginia Tech) in the queries and see if the top results change significantly.

More importantly, perhaps link structure (a by-product of authoring web pages) is too slow a metric to capture rapidly changing popular results such as team standings. In the future we plan to investigate more dynamic metrics such as the magnitude of search results or fan-based message board activity. To answer our question, although authority means quality, quality does not necessarily mean authority – at least not immediately.

6. 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] L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak. Spatial Variation in Search Engine Queries. In *Proceeding of WWW '08*, pages 357–366, 2008.

[3] R. A. Baeza-Yates, F. Saint-Jean, and C. Castillo. Web Structure, Dynamics and Page Quality. In *Proceedings of SPIRE'02*, pages 117–130, 2002.

[4] K. Bharat and G. A. Mihaila. When Experts Agree: Using Non-Affiliated Experts to Rank Popular Topics. *ACM Transactions on Information Systems*, 20(1):47–58, 2002.

[5] R. G. Capra, C. A. Lee, G. Marchionini, T. Russell, C. Shah, and F. Stutzman. Selection and Context Scoping for Digital Video Collections: an Investigation of YouTube and Blogs. In *Proceedings of JCDL '08*, pages 211–220, 2008.

[6] J. Cho and S. Roy. Impact of Search Engines on Page Popularity. In *Proceedings of WWW'04*, pages 20–29, 2004.

[7] J. Cho, S. Roy, and R. E. Adams. Page Quality: In Search of an Unbiased Web Ranking. In *Proceedings of SIGMOD'05*, pages 551–562, 2005.

[8] M. Klein and M. L. Nelson. Correlation of Music Charts and Search Engine Rankings. In *Proceedings of JCDL '09*, 2009.

[9] J. M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.

[10] E.-P. Lim, B.-Q. Vuong, H. W. Lauw, and A. Sun. Measuring Qualities of Articles Contributed by Online Communities. In *Proceedings of WI '06*, pages 81–87, 2006.

[11] F. McCown and M. L. Nelson. Agreeing to Disagree: Search Engines and Their Public Interfaces. In *JCDL '07*, pages 309–318, 2007.

[12] M. L. Nelson, M. Klein, and M. Magudamudi. Correlation of Expert and Search Engine Rankings. *CoRR*, abs/0809.2851, 2008.

[13] S. Pandey, S. Roy, C. Olston, J. Cho, and S. Chakrabarti. Shuffling a Stacked Deck: the Case for Partially Randomized Ranking of Search Engine Results. In *VLDB '05*, pages 781–792, 2005.

[14] S. Y. Rieh. Judgement of Information Quality and Cognitive Authority in the Web. *Journal of the American Society for Information Science and Technology*, 53(2):145–161, 2002.

[15] S. Y. Rieh and N. J. Belkin. Understanding Judgment of Information Quality and Cognitive Authority in the WWW. In *American Society for Information Science and Technology Annual Meeting*, pages 24–29, 1998.

[16] R. Tang, K. B. Ng, T. Strzalkowski, and P. B. Kantor. Automatically Predicting Information Quality in News Documents. In *Proceedings of NAACL'03*, pages 97–99, 2003.

[17] T. Upstill, N. Craswell, and D. Hawking. Predicting Fame and Fortune: Pagerank or Indegree? In *Proceedings of the Australasian Document Computing Symposium, ADCS2003*, pages 31–40, Canberra, Australia, December 2003.