Ranking Domain Names Using Various Rating Methods

Talattinis Kyriacos, Zervopoulou Christina, Stephanides George
Department of Applied Informatics
University of Macedonia
Thessaloniki, Greece
e-mail: ktalattinis@uom.edu.gr, zervopoulou@uom.edu.gr, steph@uom.gr

***Abstract*—This paper deals with the ranking of domain names, which is considered important because it is associated with their selling price. For this purpose, four well-known rating methods were used, the Massey method, the Colley method, the Keener method and finally, a method based on finite Markov chains and therefore, called Markov method. Although Massey's, Colley's and Keener's methods have their origin in sports teams rating, they can be modified and successfully applied for domain names ranking. Our effort to correctly rank domain names is based on search volume of the keyword of each domain and therefore, we used Google trends. We have also considered other factors, such as Alexa rank and keyword popularity. Information was collected via Internet and implementation of the models took place using computing tools. Our study is directly related to the global online information and for this reason allows us to do a more sophisticated rating model.**

Keywords-domain names; domain name ranking; rating; Massey method; Colley method; Keener method; Markov method; Kendall's tau; Google trends.

I. INTRODUCTION

This paper presents modifications of some sports teams rating methods, in order to use them in ranking of domain names. This kind of ranking is considered important because it is associated with the formation of the price at which domain names can be sold. Specifically, these two figures are proportional amounts, i.e., the higher the rank of a domain name, the higher its selling price will be.

It is a fact that with the growth of Internet, multiple sources of profitability have appeared by its use [2][9]. Thus, the concept of domain name emerged, which mainly refers to names of websites and their extensions. It has also been proved that the ownership of domain names can be particularly lucrative for their owners.

Specifically, each domain name can create value for its owner, through revenues from an active website or even without the existence of it [9]. Nowadays, due to the rapid development of e-commerce on a global level, the domain name market has already grown into a robust and profitable industry, where millions of customers search time after time for high quality domain names in order to promote their businesses. Currently, the actual value of a domain name is difficult to be accurately determined. However, there is a number of objective factors involved in determining the final selling price. The ownership of a domain name grants its owner two types of rights:

1) Managerial flexibility
2) Legal protection of trademarks.

Therefore, value can be created from a domain name in two ways: either by the expected profits or by options for action, such as the creation of an active website.

Creating an active site is not so easy, because the development of its content requires hard work, thoroughness and imagination, contrary to domain name fortification, which is achieved by a few "clicks" at the website of the pertinent regulatory authority. Acquiring a domain name has always been speculative. Aspiring investors taking advantage of new profit opportunities offered by the Internet, register a domain name and place a simple graphic like "page under construction". Then, they only have to wait for someone who has an exploitation plan for the domain name, but has however not acquired the appropriate website.

Another profitable and efficient strategy is to use a synonym for a domain name in conjunction with an intensive advertising campaign. The phenomenon of "cybersquatting" has also come of which concerns the creation of a website with a name, closely related to the name of an already popular website in order to exploit its reputation.

Domain names purchase is usually made via an auction. Investors often need to know which domain trading is the most profitable. A domain name can be considered as an investment [3] similarly to the real estate market. However, it is not clear how to estimate domain names value, because this market is relatively new. Consequently, some domain name sellers set selling prices arbitrarily without taking into account the actual value of the domain name. Domain name ranking can help investors to choose which domain to negotiate. Ranking refers only to the domain name and not to the active website.

In this paper, after Section I, which is a short introduction to the subject, follows Section II, where we discuss about some factors that determine domain names' rate. From these factors we chose the most frequently used by the majority of people involved in domain names market. Section III is an overview of the rating methods presented in the paper, while Section IV consists of some illustrative examples of how these methods can be applied in order to rank certain domain names. Finally, Section V presents the experimental results generated by these methods that refer to the ranking process of domain names. Indicatively, we present top 25 domain names as a partial list of all the domains we tested.

II. DETERMINANT FACTORS

In order to rank a group of domain names, we must first clarify which are the factors that affect their importance, their value and consequently, their rank. It is worth mentioning that this market is in embryonic stage, i.e., there is no enough literature referring to the selection criteria of these factors and no other approaches for domain name ranking have yet been proposed.

Though there are many factors that determine domain names' rank, we indicatively mention these that are usually used by the majority of people (domain traders) involved in domain names market. These factors can be easily computed and are:

1. keyword popularity: the number of search results on Google for a key-word is a good indicator of how efficient is the keyword.
2. search volume of the keyword: the comparison of keyword popularity over a period of time. Google trends is the most popular and free tool used to accomplish this task. In Google trends up to five keywords can be queried simultaneously.
3. traffic: classification of domain name in Alexa. Alexa.com is currently the most reliable counter of Internet traffic and the most popular service which publishes information on the popularity of a website. It calculates the global ranking of a domain name from the traffic it has. This calculation can be done per day, per week, per month etc. The higher the ranking is, the greater the value of the name.
4. domain name extension: the extension of a domain name, in other words, the top level domain name can affect the value and the rank of the domain name. The most dominant extension is .com. Below .com, come .net, .org and domestic extensions.
5. the size of the domain name word: Names with many characters are usually hard to memorize so those with the least possible characters are more preferred.

Some other factors that also affect domain name rank but are difficult enough to be expressed quantitatively are industry popularity and brandability. Industry popularity relates to the market volume to which a specific domain name can be applied, while brandability refers to the case that someone comes up with such an interesting new word that can become a trademark [5].

For this first approach of the subject and inspired by [12], we thought that keyword popularity, search volume of the keyword and traffic measures will have the greatest importance among five determinant factors mentioned above.

III. OVERVIEW OF METHODS

First of all, we should define what the terms rating and ranking exactly imply and realize the difference between them. Rating refers to the evaluation or assessment of an item in terms of quality, quantity or some combination of both and thus, assigns a numerical value to it. Ranking is a relationship between a set of items, i.e., for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second. Therefore, a ranking vector is a permutation of the integers 1 through n or, in other words, a sorted rating vector [1]. The methods presented in this paper are due to K. Massey, W. Colley, J.P. Keener and finally, to finite Markov chains. They are being used many years ago and had initially been invented for very different purposes. Nevertheless, they can all be used for domain names ranking or webpages ranking and more generally, for the ranking of any set of objects.

A. *Massey's Method*

This method was proposed by Kenneth Massey in 1997 for ranking college football teams. Apart from numbers of wins and losses of a team, it also considers game scores in the ratings, i.e., spread of points, via a system of linear equations [4]. Massey's method is based on the mathematical theory of least squares, which can be represented by the following equation:

$$\mathrm{r_i} - \mathrm{r_j} = \mathrm{y_k}, \tag{1}$$

where r_i and r_j are the ratings of teams i and j, respectively and y_k is the margin of victory for a game k between these teams. Each game k can be given by an equation of this form, so a system of m linear equations and n unknowns is created, where m is the number of the games that have already been played and n is the number of teams [1].This system can be written as:

$$X \cdot r = y \tag{2}$$

and is overdetermined, because $m \gg n$, i.e., there are more equations than unknowns. To deal with this problem, Massey proposed the use of matrix $M = X^T \cdot X$ instead of X, therefore, a least squares solution is obtained [7]:

$$X^T \cdot X \cdot r = X^T \cdot y \tag{3}$$

Massey matrix M can be easily filled considering that every diagonal element M_{ii} is the total number of games played by team i and every off-diagonal element M_{ij}, for $i \neq j$, is the negation of the number of games played by team i against team j. Consequently, the Massey least squares system now becomes:

$$M \cdot r = d \tag{4}$$

where $\mathrm{M_{n\times n}}$ is the Massey matrix described above,
$\mathrm{r_{n\times 1}}$ is the vector of unknown ratings and
$\mathrm{d_{n\times 1}}$ is the total difference in scores for each team.

Apart from its simple formation and its much smaller size than X, the columns of matrix M are linearly dependent, which leads to $\mathrm{rank(M)} < \mathrm{n}$ and so, the linear system $\mathrm{M} \cdot \mathrm{r} = \mathrm{d}$ does not have a unique solution [1]. Massey solved this problem by replacing any row in M with a row of all ones and the corresponding value of d with a zero. The row in M chosen by Massey is the last one.

Summarizing the Massey Rating Method, firstly we have to form the Massey matrix M and the vector d, which represents the total difference in scores for team i, then we have to force matrix M to have full rank by making some replacements and finally, we have to solve the linear system generated by these replacements in order to take ratings vector r [4]. More specifically, we can form the Massey matrix $M = X^T \cdot X$ using $M_{ij} = -n_{j,i}$, if $i \neq j$ and $M_{ij} = n_i$, if $i = j$, where n_i is the number of games played by team i and $n_{j,i}$ is the number of games played by team i against team j. The vector d of the total difference in scores for team i is given by equation $d = X^T y$. We can make the rank of matrix M full either by replacing it with $M + e^T e$, where e is a vector of all ones or by replacing one of the rows of M with e and the corresponding entry in d with c [4]. Finally, we compute the Massey rating vector r by solving the linear system generated by the previous replacement.

B. Colley's Method

This method was proposed by astrophysicist Dr. Wesley Colley in 2001 for ranking sports teams. Colley's method is based on very simple statistical principles. In fact, it is a modified form of one of the oldest rating systems, which uses the percentage of victories of each team. This percentage is given by:

$$r_i = \frac{n_w}{n_{tot}} \tag{5}$$

where n_w are the victories of group i and
n_{tot} is the total number of games played for team i [14].

Colley's method makes use of an idea from probability theory, known as Laplace's 'rule of succession', which transforms the standard winning percentage as below [14]:

$$r_i = \frac{1+n_w}{2+n_{tot}} \tag{6}$$

As follows from the above, the only information used by this model are wins, losses and number of games each team played, assuming no ties. Thus, the generated ratings are bias free, which implies that certain points gained by each team in a game are not included [4]. In other words, a win is a win regardless of the score [13]. Due to the use of Laplace's 'rule of succession', Colley's method has several advantages over the traditional rating formula:

1. At the beginning of the season, each team has a rating of $\frac{1}{2}$, instead of the preseason rating $\frac{0}{0}$ of the traditional system, which does not make any sense.
2. Colley's method takes into consideration the strength of schedule, which is the strength of a team's opponents. This implies that, if a team beats a strong opponent it ought to receive a greater reward than if it has beaten a weaker one [1].

Then follows a summary of the Colley Rating Method: At first, we can form the Colley matrix C using $C_{ij} = -n_{j,i}$, if $i \neq j$ and $C_{ij} = 2 + n_i$, if $i = j$, where n_i is the number of games played by team i and $n_{j,i}$ is the number of games played by team i against team j. Then, we compute vector b given by:

$$b_i = \frac{1+(w_i - l_i)}{2} \tag{7}$$

where w_i is the number of wins by team i and l_i is the number of loses by team i. Finally, we solve the linear system:

$$C \cdot r = b \tag{8}$$

where the r is the rating vector for the teams [4].

C. Keener's Method

This method has been proposed by James P. Keener in 1993 for football teams ranking in uneven paired competition [6]. Keener's method is based on the theory of nonnegative matrices and forms a smoothed matrix of scores [4] generated by Laplace's rule of succession:

$$\frac{S_{ij}+1}{S_{ij}+S_{ji}+2} \tag{9}$$

Laplace's rule of succession refers to computing the entry i of the Keener matrix, where S_{ij} is the points that team i scored and S_{ji} is the points scored by team j. The reason that Keener uses Laplace's rule of succession ratio is to ensure that if a team scores 0 points, the other team does not get the entirety of the points [4].

In contrast to Colley's method, Keener's method is biased, implying that a team can boost its ranking by running up its score in a game. In other words, score points do matter.

Summarizing this method, we can form Keener matrix K using:

$$K_{ij} = h\left(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}\right) \tag{10}$$

if team i played against team j, otherwise 0, where S_{ij} is number of points scored by team i against team j and

$$h(x) = \frac{1}{2} + \frac{1}{2} sgn\left(x - \frac{1}{2}\right)\sqrt{|2x-1|} \tag{11}$$

Finally, we can solve $K \cdot r = \lambda \cdot r$ to get Perron vector of matrix K, i.e., rating vector r. In the linear system given above, λ is the spectral radius (dominant eigenvalue) of K[4].

D. Markov's Method

This method utilizes finite Markov chains theory and therefore, it is called Markov Method. It was first used by graduate students, Angela Govan and Luke Ingram to successfully rank NFL football and NCAA basketball teams respectively [1], where NFL is the National Football League and NCAA is the National Collegiate Athletic Association of the United States.

Markov's method is known as Generalized Markov (GeM) ranking model and is, indeed, an adjustment of the famous PageRank algorithm that Google uses for webpage ranking. Similarly to PageRank, GeM uses parts of finite Markov chains and graph theory in order to generate ratings of n objects in a finite set. Not only sports but also any problem that can be represented as a weighted directed graph can be solved using GeM model [4].

The main idea behind the Markov Method is voting. In every game between two teams the weaker team casts a vote for the stronger team. There are many ways for a team to vote another. The simplest method uses wins and losses, implying that a winning team gains a vote by each team that has beaten. A better model would take into account game scores, namely, a winning team gets as much votes by a weaker opponent as is the margin of victory in the game between them. To make the voting method even more advanced both teams should be allowed to cast votes equal to the number of points given up in the game [1].

The main advantage of Markov's method towards the other rating methods is the combination of more than one statistics to generate rating vector r. In order to get the GeM rating vector r, we first form G using voting matrices for the p game statistics of interest [4]. This can be done by:

$$G = a_0 S_0 + \cdots + a_p S_p \tag{12}$$

where $0 \leq a_i \leq 1$ and $\sum a_i = 1$.

Each stochastic matrix S_i is called a *feature* matrix and will be formed using another statistic. Finally, we compute rating vector r, the stationary vector or dominant eigenvector of G. If G is reducible, we use the irreducible

$$\bar{G} = \beta G + (1-\beta)/\,nE,\ 0 < \beta < 1 \tag{13}$$

where E is the matrix of all ones.

IV. Illustrative Examples

Methods referred at the previous section have a wide variety of applications except of sports. Our thought was to apply these methods for ranking domain names. Therefore, we used one of the most significant determinant factors shown at Section II, which is Google trends.

Google Trends provides relative numbers. In fact, it analyzes a portion of searches done in Google in order to compute how many of them have been done for the terms entered, compared to the total number of searches done on Google over time. Google does not reveal absolute numbers for competitive reasons, but also because those numbers would not be exact. The fact that Google trends are relative numbers implies that there may have been done more searches for object A than for object B, but these searches may be less than those of another object C. For example, assuming that object A is Gauss and object B is Markov, the winner is Gauss with 54 Google trends points average against Markov's 27 points average. However, if object C is Shannon, Gauss becomes the underdog with 9 points average, while Shannon is given 54. This much similarity between Google trends and points in a game is exactly the reason we decided to use Google trends as determinant factor for the ranking methods presented here.

In the example described below, there are five domain names that have been sold in early 2014, which are jean.com, desirous.com, authorization.com, true.com and finally, peaked.com. We will attempt to rank these domains based on search volume average they get by Google trends during 2013. The question is how can search volume average be related to the points that a team succeeded against another?

There are many ways to define the notion of a game for domain names. For example, if statistics on domain names are given by Google trends, then we can say that domain i beats domain j if $d_i > d_j$, where d_i and d_j are the Google trends measures for these domains. Therefore, $d_i - d_j$ represents the difference in trends' value between domains i and j. Table I shows Google trends data for the five domains of our example.

TABLE I. GOOGLE TRENDS DATA

Domain i	**Domain j**	**Trends i, j**
jean.com	desirous.com	88, 0
jean.com	authorization.com	88, 4
jean.com	true.com	76, 73
jean.com	peaked.com	88, 0
desirous.com	authorization.com	1, 93
desirous.com	true.com	0, 73
desirous.com	peaked.com	20, 80
authorization.com	true.com	3, 73
authorization.com	peaked.com	93, 5
true.com	peaked.com	73, 0

Adjusting the Massey rating method for domain names, we start with the same idealized function (1). Then, the Massey domain ranking method proceeds as usual, according to (4). Below this linear system is showed:

$$\begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} 263 \\ -313 \\ 26 \\ 213 \\ 0 \end{pmatrix} \tag{14}$$

Table II gives rating and ranking data generated by Massey method:

TABLE II. RATING AND RANKING BY MASSEY METHOD

Ranking	**Domain**	**Rating**
1	jean.com	52.6
2	true.com	42.6
3	authorization.com	5.2
4	peaked.com	-37.8
5	desirous.com	-62.6

As we conclude from the above matrix, jean.com has beaten all the other four domain names and, thus, it terminates at first position of ranking. Contrary to jean.com, domain desirous.com has been defeated by all others,

 ISBN: 978-1-61208-346-9

therefore, it has the lowest rating of all and so, it takes the last position of ranking.

Using the Colley rating method (8) for domain names, we get the results below:

$$\begin{pmatrix} 6 & -1 & -1 & -1 & -1 \\ -1 & 6 & -1 & -1 & -1 \\ -1 & -1 & 6 & -1 & -1 \\ -1 & -1 & -1 & 6 & -1 \\ -1 & -1 & -1 & -1 & 6 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 1 \\ 2 \\ 0 \end{pmatrix} \tag{15}$$

Table III gives rating and ranking data generated by Colley method. Rating values have been rounded at four decimal digits.

TABLE III. RATING AND RANKING BY COLLEY METHOD

Ranking	Domain	Rating
1	jean.com	0.7857
2	true.com	0.6429
3	authorization.com	0.5
4	peaked.com	0.3571
5	desirous.com	0.2143

As we may see in the above table, jean.com terminates again first, while desirous.com gets again the last rank. Positions of the three other domains also remain the same.

Then, we continue with the Keener rating method for domain names. Below, the values of Keener matrix K are shown:

$$\begin{pmatrix} 0 & 0.9944 & 0.9727 & 0.5705 & 0.9944 \\ 0.0056 & 0 & 0.0105 & 0.0067 & 0.1165 \\ 0.0273 & 0.9895 & 0 & 0.0263 & 0.969 \\ 0.4295 & 0.9933 & 0.9737 & 0 & 0.9933 \\ 0.0056 & 0.8835 & 0.031 & 0.0067 & 0 \end{pmatrix} \tag{16}$$

Table IV gives rating and ranking data generated by Keener method:

TABLE IV. RATING AND RANKING BY KEENER METHOD

Ranking	Domain	Rating
1	jean.com	0.7391
2	true.com	0.6604
3	authorization.com	0.1253
4	peaked.com	0.039
5	desirous.com	0.0192

Table IV indicates more significant difference between true.com and authorization.com than before. This can be explained by the fact that Keener's method is not bias-free, which means that all the points succeeded in a duel are taken into account for ranking.

The last method we used in order to rank the five domain names of our example is Markov method. As we have mentioned in Section II, the main idea behind this method is voting. In Table V, follows the trends voting matrix:

TABLE V. TRENDS VOTING MATRIX

	jean	desirous	authorization	true	peaked
jean	0	0	4	73	0
desirous	88	0	93	73	80
authorization	88	1	0	73	5
true	76	0	3	0	0
peaked	88	20	93	73	0

Below, stochastic matrix G is shown, which is generated by normalizing the rows of the above voting matrix:

$$\begin{pmatrix} 0 & 0 & 0.0519 & 0.9481 & 0 \\ 0.2635 & 0 & 0.2784 & 0.2186 & 0.2395 \\ 0.5269 & 0.006 & 0 & 0.4371 & 0.03 \\ 0.962 & 0 & 0.038 & 0 & 0 \\ 0.3212 & 0.073 & 0.3394 & 0.2664 & 0 \end{pmatrix} \tag{17}$$

Table VI gives rating and ranking data generated by Markov method:

TABLE VI. RATING AND RANKING BY MARKOV METHOD

Ranking	Domain	Rating
1	jean.com	0.4801
2	true.com	0.4746
3	authorization.com	0.0435
4	peaked.com	0.0014
5	desirous.com	0.0004

In the small example described in this paper, all four methods generate same ranking results. However, Markov's method has a vital difference of the other three. This difference comes from the fact that, as we have mentioned in Section III, Markov method allows the use of more than one statistics. Therefore, Markov method can be characterized as more representative than the others. The weights we have set for Google trends, Google results and Alexa rank were 0.4, 0.3 and 0.3 respectively, due to our intention to rely mostly on Google trends, though these weights may vary in all the possible ways.

At this point, we will see the results generated by some other determinant factors, such as Google Result and Alexa Rank.

TABLE VII. OTHER DETERMINANT FACTORS

Domain	Google Result	Alexa Rank (ar)	Adjusted Alexa Score (ads)
jean.com	358,000,000	1,782,928	560.8751
desirous.com	1,660,000	-	1
authorization.com	72,700,000	3,566,076	280.4203
true.com	676,000,000	220,338	4538.4818
peaked.com	5,780,000	-	1

As we have mentioned in Section II, Alexa classifies domain names counting the Internet traffic. Alexa ranks only domain names with an active website, thus dashes shown in Table VII represent domain names with a non-active website. At this point, we should deal with a major issue, which is that though Alexa is a ranking system, we want to turn it into one that uses points.

This issue implies that, though Alexa assigns to rank "1" the most visited website, supposedly A, this value as a number is smaller than ranks of much less visited websites, supposedly B, C, etc. Therefore, we should modify Alexa rankings given in Table VII, so that the ordering of the ranking values is reversed. One solution to this problem is, for n items ranked by Alexa, to compute each item's rank by:

$$ads_i = \frac{\max\{ar_i, \dots, ar_n\}+1}{ar_i} \tag{18}$$

 ISBN: 978-1-61208-346-9

where ads_i is the adjusted Alexa rank for website i and ar_i is the rank given by Alexa for website i. However, this solution is not so fair for some items of the set. For instance, in case we have to rank domains google.com, facebook.com and desirous.com, rank given by Alexa is “1”, “2” and “no enough data to rank this website”, respectively. Thus, when dividing the total number of items ranked, namely “3”, with each item’s rank we get the results below:

- Three points are assigned to google.com
- Two points is assigned to facebook.com and
- One point is also assigned to desirous.com

The unfairness of this solution lies in that points of facebook.com, which is an active website, are very close to points of a non-active website, as is desirous.com. Thus, we should use a better solution, which is described below.

This better solution might be obtained using the following equation:

$$ads_i = \frac{tna}{ar_i} \quad (19)$$

where tna is the total number of websites ranked by Alexa. Despite of our thorough research, we have not found any official source referring exactly how many websites are currently ranked by Alexa. Therefore, we have chosen a typically large number, namely, $1 \cdot 10^9$ for tna variable mentioned before, i.e., we divide $1 \cdot 10^9$ with each item’s rank to turn the ordering of the ranking values into descending. In Table VII, we have written the adjusted Alexa score value of each domain.

As we can see in Table VII, true.com has the higher traffic amount of all, while before adjustment it was at the third position of Alexa rank. Then comes jean.com and authorization.com, which is the domain showed up to be at the first position before adjustment. Domains desirous.com and peaked.com take both the value “1” for their traffic, which is the value assigned to non-active websites.

Finally, in Table VIII, we tested the five domains of our example applying Markov method with three determinant factors, which are Google trends, Google results and Alexa rank. The results generated are shown in the table below.

TABLE VIII. MARKOV RATING - RANKING WITH 3 STATISTICS

Ranking	Domain	Rating
1	true.com	0.4649
2	jean.com	0.4242
3	authorization.com	0.1057
4	peaked.com	0.0042
5	desirous.com	0.0011

Table IX is a summary table, which shows the domain rankings generated by all four methods described in Section III and their selling prices:

TABLE IX. SUMMARY TABLE

Domain	Massey	Colley	Keener	Markov 1	Markov 3	Prices
jean	1	1	1	1	2	50000
desirous	5	5	5	5	5	2600
author ization	3	3	3	3	3	35100
true	2	2	2	2	1	350000
peaked	4	4	4	4	4	4000

As we see, due to the use of more than one determinant factors in this ranking, rank positions between the first two domains have interchanged and this agrees with the selling prices.

V. EXPERIMENTAL RESULTS

At first, we should describe our database, which contains information on transaction prices collected from publicly available tenders and values from databases of domain names coming from closed auctions data. Currently the database consists of 75,000 transaction prices that occurred during the period between 1999 and 2013. The database is updated regularly and it is worth mentioning that the collection of data required some effort since selling prices are not always available in digital form, even when they are published. In order to gather our data from Internet resources, we have implemented a web crawler in Java. Our crawler is typically programmed to visit sites, which contain domain names’ selling prices, Google trends, keyword popularity and traffic measures. The crawling process was held by taking into account the reliability of information. The data gathering process also involved parsing data files.

In any case, data gathering and parsing had to be automated. We have conducted thorough research and have already implemented some techniques for parallelization of collecting data, in order to keep our database updated in time. For more details about crawling, its parallelization and parsing processes we refer the reader to [8].

The rest of this paper presents the empirical results generated by the four methods described at Section III when we applied them to our database data. The numerical computations of the ratings were done using Matlab. Tables XI to XIV have been constructed in the following format: the first column represents the ranking of domains, the second column is the domain name itself, the third column represents the rating of each domain, the fourth column is the price at which the domain name was sold and finally, the last column consists of the date on which each domain was sold. Though we refer to selling price, it cannot be a reliable measure of comparison, due to its dependance of the time that happened.

In Tables XI to XIV, we present top 25 domain names as a partial list of all the domains we tested, using Google trends determinant factor for the year 2013. Each table shows the top 25 domains, according to one of the methods. We tested domain names with same Top Level Domain (TLD), i.e., .com. Table XIV shows the results of Markov method using three determinant factors, Google trends, Google results and Alexa rank. Similarly to Section IV, the weights we used for these factors were 0.4, 0.3 and 0.3 respectively. Due to the reliability of method Markov that takes into account three determinant factors, we have also included in its table the importance degree given to the domain by Google PageRank (PR). Briefly, PageRank algorithm states that a website is important if it is shown by other important websites. This degree gets values from 1 to 10 (PR1 - PR10). The higher the PageRank obtained from a website, the higher its ranking position in search results [11]. The comparison among Markov method and Google PageRank shows that there are many results in common between them.

In order to compare the generated ranking lists, we make use of Kendall 's correlation measure τ, which gives the degree to which one list agrees (or disagrees) with another [1] and is computed by:

Kendall' s tau $\tau = \frac{n_c - n_d}{n(n-1)/2}$ (19), where n_c is the number of concordant pairs and n_d is the number of discordant pairs.

Kendall 's tau value varies between -1 and 1, i.e., $-1 \leq \tau \leq 1$. If $\tau = 1$, then the two lists are in perfect agreement, while if $\tau = -1$, the two lists are totally opposite to each other [10]. Comparing the methods described in this paper, according to Kendall 's tau, we get the results below:

TABLE X. KENDALL'S TAU TABLE

Pair of Methods	Kendall's Tau Value
Massey - Colley	0.942
Massey - Keener	0.9451
Massey - Markov	-0.451
Colley - Keener	0.9969
Colley - Markov	-0.4431
Keener - Markov	-0.44

From Table X, we conclude that methods Massey, Colley and Keener are very alike, while Markov is differentiated due to the use of more than one determinant factors that provides.

Then, follow the tables that show the top 25 domain names according to each of the four methods we described.

TABLE XI. MASSEY RANKING

Ranking	Domain name	Rating	Price	Selling Date
1	fb.com	78.6078	8,500,000	1/1/2010
2	phone.com	69.8627	1,200,000	1/2/2003
3	shop.com	61.2745	3,500,000	1/11/2003
4	photo.com	58.7451	1,250,000	6/5/2010
5	men.com	54.3333	1,320,000	1/2/2000
6	software.com	52.5882	3,200,000	1/12/2005
7	find.com	51.2549	1,200,000	1/3/2004
8	pizza.com	51.1176	2,605,000	3/4/2008
9	express.com	41.8235	2,000,000	1/3/2000
10	call.com	38.6275	1,100,000	2/9/2009
11	tom.com	35.9608	2,500,000	1/12/1999
12	zip.com	34.5294	1,058,830	28/10/2010
13	candy.com	28.6078	3,000,000	10/6/2009
14	vista.com	21.0196	1,250,000	14/11/2007
15	ticket.com	20.5686	1,525,000	16/10/2009
16	coupons.com	20.2746	2,200,000	1/1/2000
17	fly.com	19.9804	1,500,000	1/11/1999
18	wine.com	16.5882	3,300,000	1/9/2003
19	webcam.com	15.3725	1,020,000	10/6/2009
20	beer.com	13.9216	7,000,000	1/1/2004
21	england.com	12.9412	2,000,000	1/12/1999
22	casino.com	12.1961	5,500,000	1/11/2003
23	telephone.com	9.6275	2,000,000	1/1/2000
24	VIP.com	8.5294	1,400,000	1/12/2003
25	autos.com	8.2746	2,200,000	1/12/1999

TABLE XII. COLLEY RANKING

Ranking	Domain name	Rating	Price	Selling Date
1	fb.com	0.9717	8,500,000	1/1/2010
2	phone.com	0.9528	1,200,000	1/2/2003
3	shop.com	0.934	3,500,000	1/11/2003
4	photo.com	0.9151	1,250,000	6/5/2010
5	pizza.com	0.8962	2,605,000	3/4/2008
6	men.com	0.8774	1,320,000	1/2/2000
7	express.com	0.8585	2,000,000	1/3/2000
8	call.com	0.8302	1,100,000	2/9/2009
9	software.com	0.8302	3,200,000	1/12/2005
10	find.com	0.8019	1,200,000	1/3/2004
11	tom.com	0.783	2,500,000	1/12/1999
12	candy.com	0.7641	3,000,000	10/6/2009
13	zip.com	0.7453	1,058,830	28/10/2010
14	coupons.com	0.7264	2,200,000	1/1/2000
15	wine.com	0.7075	3,300,000	1/9/2003
16	ticket.com	0.6887	1,525,000	16/10/2009
17	webcam.com	0.6698	1,020,000	10/6/2009
18	england.com	0.6509	2,000,000	1/12/1999
19	vista.com	0.6321	1,250,000	14/11/2007
20	fly.com	0.6132	1,500,000	1/11/1999
21	korea.com	0.5849	5,000,000	1/1/2000
22	beer.com	0.5849	7,000,000	1/1/2004
23	casino.com	0.5566	5,500,000	1/11/2003
24	telephone.com	0.5377	2,000,000	1/1/2000
25	VIP.com	0.5189	1,400,000	1/12/2003

TABLE XIII. KEENER RANKING

Ranking	Domain name	Rating	Price	Selling Date
1	fb.com	0.3224	8,500,000	1/1/2010
2	phone.com	0.2813	1,200,000	1/2/2003
3	shop.com	0.2685	3,500,000	1/11/2003
4	photo.com	0.2547	1,250,000	6/5/2010
5	pizza.com	0.246	2,605,000	3/4/2008
6	men.com	0.2422	1,320,000	1/2/2000
7	express.com	0.2378	2,000,000	1/3/2000
8	software.com	0.2196	3,200,000	1/12/2005
9	call.com	0.2194	1,100,000	2/9/2009
10	find.com	0.2137	1,200,000	1/3/2004
11	tom.com	0.1964	2,500,000	1/12/1999
12	candy.com	0.1771	3,000,000	10/6/2009
13	zip.com	0.1716	1,058,830	28/10/2010
14	coupons.com	0.152	2,200,000	1/1/2000
15	wine.com	0.1429	3,300,000	1/9/2003
16	ticket.com	0.1391	1,525,000	16/10/2009
17	webcam.com	0.1374	1,020,000	10/6/2009
18	england.com	0.133	2,000,000	1/12/1999
19	vista.com	0.1313	1,250,000	14/11/2007
20	fly.com	0.128	1,500,000	1/11/1999
21	korea.com	0.1169	5,000,000	1/1/2000
22	beer.com	0.1158	7,000,000	1/1/2004
23	casino.com	0.1127	5,500,000	1/11/2003
24	telephone.com	0.1014	2,000,000	1/1/2000
25	VIP.com	0.1001	1,400,000	1/12/2003

TABLE XIV. MARKOV RANKING WITH 3 STATISTICS

#	Domain name	Rating	Google PageRank	Price / Selling Date	
1	coupons.com	0.1055	6	2,200,000	1/1/2000
2	photo.com	0.1053	4	1,250,000	6/5/2010
3	shop.com	0.0816	5	3,500,000	1/11/2003
4	VIP.com	0.0539	3	1,400,000	1/12/2003
5	find.com	0.0472	4	1,200,000	1/3/2004

6	casino.com	0.046	5	5,500,000	1/11/2003
7	phone.com	0.0416	5	1,200,000	1/2/2003
8	express.com	0.0383	5	2,000,000	1/3/2000
9	fb.com	0.0346	0	8,500,000	1/1/2010
10	men.com	0.034	3	1,320,000	1/2/2000
11	tom.com	0.0306	7	2,500,000	1/12/1999
12	software.com	0.0288	5	3,200,000	1/12/2005
13	call.com	0.0268	4	1,100,000	2/9/2009
14	pizza.com	0.0217	4	2,605,000	3/4/2008
15	feedback.com	0.0199	7	1,230,000	1/2/2003
16	zip.com	0.018	0	1,058,830	28/10/2010
17	savings.com	0.0172	5	1,900,000	1/2/2003
18	wine.com	0.0166	6	3,300,000	1/9/2003
19	fly.com	0.0163	5	1,500,000	1/11/1999
20	candy.com	0.0149	5	3,000,000	10/6/2009
21	vista.com	0.0146	0	1,250,000	14/11/2007
22	webcam.com	0.0132	0	1,020,000	10/6/2009
23	england.com	0.0127	2	2,000,000	1/12/1999
24	auction.com	0.0122	5	1,700,000	27/3/2009
25	ticket.com	0.0122	0	1,525,000	16/10/2009

VI. CONCLUSIONS – FUTURE PLANS

In this paper, we saw how we can rank domain names with four different methods. These methods are mainly used in sports industry. Compared to the others, Markov method allows ranking based on more than just one factor and as we saw in Section IV, this is crucial in the formation of ranking values.

For generating our empirical results using Massey, Colley and Keener method, we used Google trends as determinant factor, while in Markov method, we used Google trends, Google results and Alexa rank. Determinant factors were chosen according to what are people involved in domain name market looking for. In our empirical results, we cannot use the selling price as criterion to check if rankings lists generated via different ranking methods match. This is due to the fact that selling prices were formed based on past factors or data, while our ranking is based on current factors or data. This is also confirmed by PageRank indicator.

In conclusion, rating methods presented in the paper may be used by many groups of people, such as domain traders, portfolio managers and investors. Concerning to decision making process, i.e., if someone decides to purchase a domain name according to its rank, the methods presented in this paper can be a utility tool, but not the only one.

Finally, when we compared the ranking lists according to Kendall's tau correlation method, we conclude that Massey, Colley and Keener have much in common, while Markov is different enough due to the factors it takes into account.

Our goal for the future is to use more determinant factors for domain names ranking, such as brandability and internet popularity, but also to test even more rating methods, such as Elo's system or the Park-Newman method.

REFERENCES

[1] A. N. Langville and C. D. Meyer, Who's #1?: The Science of Rating and Ranking, Princeton University Press, Princeton, NJ, USA, 2012.

[2] A. Tajirian, "Thoughts on Domain Name Investing for Newbies", 2008 [Online]. Available from: http://domainmart.com/news/Thoughts_on_Domain_Name_Investing_for_Newbies.pdf, [retrieved: May, 2014].

[3] A. Tajirian, "Valuing Domain Names: Methodology", 2005 [Online]. Available from: http://domainmart.com/news/methodology.htm [retrieved: May, 2014].

[4] A.Y. Govan, "Ranking Theory with Application to Popular Sports", unpublished. PhD thesis, North Carolina State University, 2008.

[5] http://ezinearticles.com/?Domain-Appraisal-Guide---20-Factors-That-Decide-the-Selling-Price&id=1436181 [retrieved: May, 2014].

[6] J. P. Keener, "The Perron-Frobenius Theorem and the Ranking of Football Teams". SIAM Review, vol. 35, No 1, March. 1993, pp. 80-93.

[7] K. Massey, "Statistical Models Applied to the Rating of Sports Teams", unpublished. Bachelor's thesis, Bluefield College, 1997.

[8] K.Talattinis, A Sidiropoulou, K.Chalkias, and G.Stephanides, "Parallel Collection of Live Data Using Hadoop", IEEE 14th PanHellenic Conference on Informatics (PCI), ISBN: 978-1-4244-7838-5, Sept. 2010, pp. 66-71.

[9] M. Jindra, "The market for Internet domain names", in Proc. 16th ITS Regional Conf., Porto, Portugal, 2005 [Online]. Available from: http://userpage.fu-berlin.de/~jmueller/its/conf/porto05/papers/Jindra.pdf [retrieved: May, 2014].

[10] M. Kendall, "A new measure of rank correlation", Biometrica, 1938.

[11] S. Brin, L. Page, R. Motwami, and T. Winograd, The PageRank citation ranking: Bringing order to the Web. Technical Report 1999 – 0120, Computer Science Department, Stanford University, 1999.

[12] T. Preis, H. S. Moat, and H. E. Stanley, Quantifying trading behavior in financial markets using Google Trends. Sci. Rep. 3, 1684, 2013.

[13] T.P. Chartier, E. Kreutzer, A.N. Langville, and K.E. Pedings, "Bracketology: How can math help?". In: Gallian, Joseph (Ed.), Mathematics and Sports Dolcani Mathematical Expositions, vol. 43. Mathematical Association of America. Jan. 2010, pp. 55-70.

[14] W.N.Colley, "Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained", 2002 [Online]. Available from: http://www.colleyrankings.com [retrieved: May, 2014].