HOW TO EXPLORE PUBLIC INTEREST FOR SPORTS? NEW METHODOLOGICAL APPROACH

Dražen Čular[1,2,3], Josip Miočić[4], Igor Jelaska[1]

[1]University of Split, Faculty of Kinesiology, Split, Croatia
[2]Croatian Institute for Kinesiology & Sport, Split, Croatia
[3]Einstein, Craft for Research, Development, Education, Trade & Service, Split, Croatia
[4]Department of Culture and Communication Sciences, University of Zadar

Original scientific paper

Abstract

Although government bodies, sport associations and funding organizations often need exact criterion for assessment of interests and trends for specific sports, there is no single definition of popularity determinants, as well as specific tool designed for reliable and valid measurement of particular sport's popularity. Consequently, there is lack of scientific-based evidence about the public interests in particular sport. This paper introduces the *Google Trends* (GT) as a web-based tool to identify and compare public interests for 5 Olympic team sports (volleyball, handball, water polo, football, basketball) according to the worldwide public Google queries transformed into 1-100 Likert type scale. Observing last 5 years using GT, it can be concluded that football (mean = 30) is the most popular analyzed sport followed by basketball (mean = 8), volleyball (mean = 3), handball (mean =1) and water polo (mean ≤1). It can be concluded that, although limited, GT can provide valuable information about public interests for specific sport on the global, regional, or local level.

***Key words**: sport popularity, tool for measuring interest for sport, big data, sport categorization, football, basketball, volleyball, handball, water polo*

Introduction

It is very difficult to determine what the world's most popular sports are. In addition, there is a problem with valid and reliable identification of the world's most popular sports, because there is not a clear definition of sport's popularity that has been applied and widely accepted. There are very few studies based on a bottom-up approach, i.e., striving to establish a framework describing public interest, whatever this interest might be (Schummer, 2005; Baram-Tsabari et al., 2008; 2009; Falchetti et al., 2007). Google trends (GT) keywords entered into search engines can be a useful resource for detecting people's information needs (Murata, 2006; Segev & Ahituv, 2010). This approach has been used to study different trends and interests worldwide, but, according to the available information, has not been used yet for measuring specific sports interest. According to Nuti et al. (2014), GT provides access to Internet search patterns by analyzing a portion of all web queries on the Google Search website and other affiliated Google sites. Beside curiosity, why is it important to have a precise tool to compare and measure interest for specific sports? On the global scene, the Olympic Games (OG) organization conducts a selection process for including sports in the summer or winter OG program. According to the Encyclopedia Britannica (https://www.britannica.com/story/how-are-sports-chosen-for-the-olympics), when choosing sports to include in the OG program, the IOC, beside other indicators, must take into consideration public interest. The Tokyo 2020 IOC committee listed the following criteria for new sports induction; added value; youth appeal; attractiveness for TV, media and the general public; gender equality; minimum impact on the number of events and/or quotas, infrastructure and operational costs and complexity (https://www.topendsports.com/events/summer/sport-selection.htm). European funds, as per example Erasmus+ sport fund, support sports projects. It is expensive and nonpractical to run research for each project to measure the impact of the project activities on interest for some sport or recreation activities. On the regional scene, government bodies and local funding organizations must make the decision for funding according to some indicators, and one of the criteria is often popularity/interest for some activity or sport. The National Sports Program (NPS 2018), as the main Croatian strategic sport document adopted by the Croatian Parliament, defines categorization of sports as special aim 1.1. National sports organizations need cost & benefit analyses before organizing sports events. Interest for specific sport can also be a valuable indicator for Sport talents detection policy makers. Is the GT tool one of the possible solutions for all above mensoned

questiones closely conected with popularity or interest for specific sport ?

Methods

Our Research uses a method of data mining (knowledge discovery in databases), which can be used both as bottom-up and top-down approach to discovering patterns (Liu and Ruiz, 2008). An example of a bottom-up approach in this research context is finding the most popular queries related to five specific Olympic team sports. In Google Trends web page (https://trends.google.com/trends/?geo=US) we inserted the following query: the names of five sports (football, volleyball, handball, basketball, water polo), & level: Worldwide & time frame: past 5 years (November 1, 2015 - October 31, 2020) & category: sports & type: web search. Data were downloaded data in .cvs format, using Microsoft Excel for Mac application version 16.43. Average score (mean) for each sport was calculated and took as screen shot of figures with graphical data from the GT web page. Values are calculated on a scale from 0 to 100, where 100 is the location with the most popularity as a fraction of total searches, per example a value of 50 indicates half as popular in comparison with the first-ranked sport. A value of <1 indicates there was not enough data for calculating the score for this term. It is important to underline that a higher value means a higher proportion of all queries, not a higher absolute query count. For comparison of GT results we also extracted data using the Toolforge website (https://pageviews.toolforge.org/?project=en.wikipedia.org&platform=all access&agent=user&redirects=0&range=latest-20&pages=), which measures popularity or interest according to page traffic data for Wikipedia sports web pages for the same **period**.

Results and discussion

Within table 1, numbers represent search interest relative to the highest point on the chart for the given region and time. Values are calculated on a scale from 0 to 100, where 100 is the location with the most popularity as a fraction of total searches in that location. A value of <1 indicates a location where there was not enough data for particular term. **Note:** A higher value means a higher proportion of all queries, not a higher absolute query count.

Table 1. Average score (mean) of interest (period from November 1, 2015 to October 31, 2020)

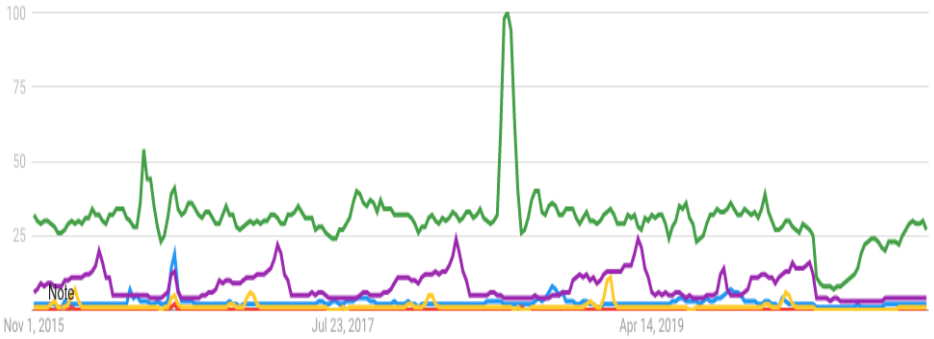
Sport	Average Score	Percentage of Max
Football	30	100 %
Basketball	8	26.7 %
Volleyball	3	10.0 %
Handball	1	3.3 %
Water polo	<1*	<1 %*

Data source: Google Trends (https://www.google.com/trends).

*Water polo does not have enough query (<1) to be calculated

Results presented in Table 1. show that football (mean = 30) is without any doubt the most popular sport according to the Google Trends queries in the period of the last 5 years, (from November 1, 2015 to October 31, 2020) Other four sports according to the queries are ranked as: 2nd basketball (mean= 8); 3rd volleyball (mean= 3); 4th handball (mean = 1); 5th water polo (mean= <1).

Figure 1. GT Trend of interest over time for the last 5 years (period from November 1, 2015 to October 31, 2020)



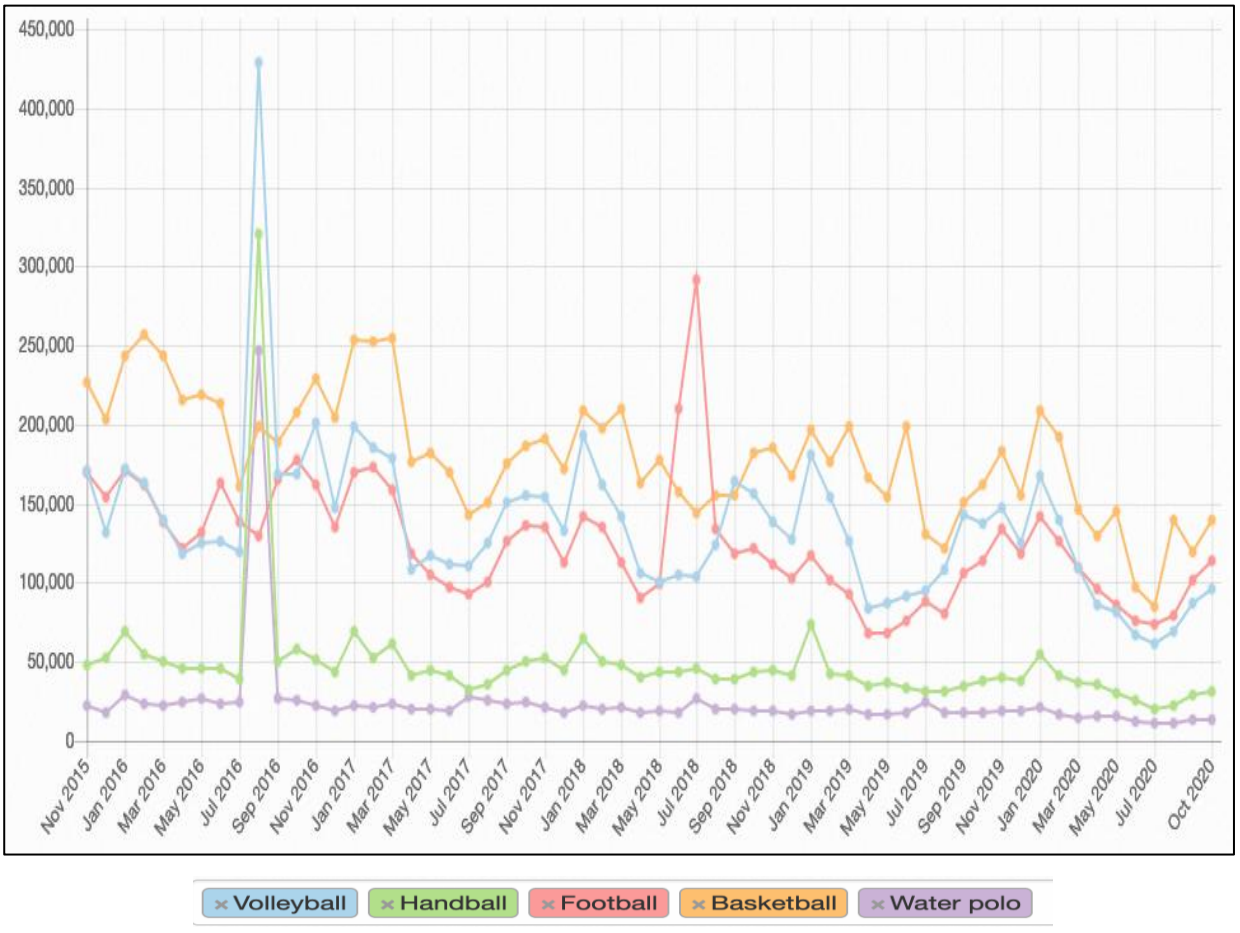
Data source: Google Trends (https://www.google.com/trends).

Legend: Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the sport. A score of <1 means there was not enough data for this term.

Results presented in Figure 1. show change of interest overtime. Certain peaks of interest can be noticed in each sport in the analyzed period. The biggest peak of interest for football was in June 2018 during the FIFA World Cup in Russia, as expected. Another peak of interest for football was in June 2016, in the period of the FIFA World Cup held in Japan. Interest for basketball had 4 Peaks, in March of each year. All Peaks are correlated with the Finals of the USA NBA competition. The interest for handball rose in January of each analyzed year, which is correlated with the period of the World/European Men's Handball Championships. The best top interest for handball was in January 2019 during the IHF World Men's Handball Championship. Interest for volleyball, handball and water polo rose in August 2016 in the period of the Olympic Games. In that period, volleyball attracted most queries of all the analyzed sports, besides football, and it is interesting that it was the only point when volleyball attracted more interest than basketball. In Table 2, the trend and comparison of Wikipedia sport web page view s trends (period from November 1, 2015 to October 31, 2020) is presented. It is interesting that mean peaks of interest are similar to the Google trends results (2016 Olympic Games, 2018 FIFA World Cup, EHF – Word/European Handball Championships (2016/ 17/18/19/20), etc. Volleyball attracted the most interest in the period of the 2016 OG and football is not so dominant except in the period of the 2018 World Cup in Russia. Handball and water polo definitely attracted less interest than other three sports.

Figure 2. Comparison of Wikipedia sport web page views trends

(period from November 1, 2015 to October 31, 2020)



Data source: (https://pageviews.toolforge.org/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&range=latest-20&pages=).

Table 2. Interest by top 5 countries (period from November 1, 2015 to October 31, 2020)

Volleyball (%)	Water polo (%)	Handball (%)	Football (%)	Basketball (%)
Thailand 100	Montenegro 100	Denmark 100	Thailand 100	Lithuania 100
Italy 70	Hungary 54	Norway 72	Laos 49	United states 56
Poland 65	Serbia 52	Croatia 60	Uganda 41	Serbia 35
Iran 58	Croatia 38	North Macedonia 54	Kenya 23	Greece 33
Philippines 58	Malta 31	Germany 49	Congo 23	Dominic Rep 32
57 countries	59 countries	55 countries	71 countries	48 countries

Data source: Google Trends (https://www.google.com/trends).

Legend: Values are calculated on a scale from 0 to 100, where 100 is the location with the most popularity as a fraction of total searches in that location, whereas a value of 50 indicates a location which is half as popular. A value of <1 indicates a location where there was not enough data for this term. **Note:** A higher value means a higher proportion of all queries, not a higher absolute query count

In Table 2., interest for each sport in top 5 countries is presented. According to the google queries in the selected period, volleyball is the most popular sport in Thailand, water polo in Montenegro, handball in Denmark, football in Thailand, and basketball in Lithuania. It is interesting that all top 5 Cities with interest for Handball and Water polo are form Europe.

Table 3. Interest by top5 cities (period from November 1, 2015 to October 31, 2020)

	Volleyball (%)	**Water polo** (%)	**Handball** (%)	**Football** (%)	**Basketball** (%)
1.	Bangkok 100	Budapest 100	Copenhagen 100	Laem Chabang 100	Louisville 100
2.	Milan 78	Belgrade 89	Oslo 80	Bangkok 69	Raleigh 65
3.	Quezon City 71	Split 80	Drochtersen 80	Nairobi 17	Omaha 60
4.	Warsaw 60	Genoa 68	Zagreb 74	Lagos 11	Indianapolis 55
5.	Rio de Janeiro 60	Zagreb 55	Stuttgart 68	Ho Chi Minh C 7	Charlotte 42

Data source: Google Trends (https://www.google.com/trends).

In Table 3., interest for each sport in top 5 cityes is presented. According to the google queries in the selected period, volleyball is the most popular sport in Bangkok (Thailand), water polo in Budapest (Hungary), handball in Copenhagen (Denmark), football in Leam Chaban (Thailand), and basketball in Louisville (USA).

Table 4. Average monthly Wikipedia sports web page view (period from 2016 to 2019)

Sport	Average monthly view
Football	6588
Basketball	6156
Volleyball	4732
Handball	/
Water polo	/

Data source: Top end Sports web page (https://www.topendsports.com/world/lists/popular-sport/wiki-sports.htm).

It is interesting that our research results are similar to the ranking of sports interest that we found on Topend Sports web page (presented in Table 4.), which is presented according to the page traffic data for Wikipedia sports pages in a four-year period (2016-2019). In the table consisting of 25 most popular sports web pages we did not find handball and water polo data.

Discussion

The most watched, the most played, or the most talked about are just some of the factors which could help determine the lists of popular sports, include the number of spectators, the number of national associations, the number of registered athletes, tournaments held, etc. Many of these should be factors to determine top sports, though it would be difficult to combine them into a single list. According to the web portal Top end Sports *(https://www.topendsports.com/world/lists/popular-sport/most-viewed.htm)*, when discussing this question, there is a difference between the *most viewed* sports (e.g., most fans) versus the *most played* sports around the world. One of the methods used is also the analysis of internet traffic to the sports sections of the IOC website (which uses data from the Olympic Program Commission Report) Wong (2012). The data presented is the number of

visits to the sport's pages of the IOC website in a certain period. On the Top end Sports Web page we found a list of sports popularity/interest according to the page traffic data for Wikipedia sports pages for a four-year period from the start of 2016 to the end of 2019, extracted using the Toolforge website (https://pageviews.toolforge.org/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&range=latest-20&pages=).

This method also has some limitations: it only notes the most popular pages - the number of page visitors can depend on the quality of the page information, search engine rankings, and strength of competing websites. It also does not take into consideration internet traffic to other pages for that sport, e.g., accessing the Wikipedia Super Bowl page does not affect the American Football page stats. Also, Wikipedia uses page redirects, for example, Soccer redirects to Association Football, and the page originally accessed is counted so stats are split between each page. Only English pages were analyzed, so the data is biased to those with internet access in the English-speaking world. To describe public interest in different areas, researchers use a top-down and a bottom-up approach. According to Baram-Tsabari (2011), most popularity lists are based on a top-down approach, i.e., identifying specific concepts that should be interesting or important to the public according to experts. Kovacic at all (2017) study aims to determine whether there are needs for athlets education, when the time for engaging in a secondary career comes. Google Trends can be useful tool for detecting education and dual carrier specific interest wich can help goverment bodyes and sport organization as policy makers. A Google Trends description of the user interface is shown on the web page (https://support.google.com/trends/). Users are able to download the output of their searches to conduct further analyses. The portal determines the proportion of searches for a user-specified term among all searches performed on Google Search. It then provides a relative search volume (RSV), which is the query share of a particular term for a given location and time period, normalized by the highest query share of that term over the time-series. Mondria, Wu. (2013) and Choi, Varian. (2012). The user can specify the geographic area to study, whether a city, country, or the world; data is available for all countries worldwide. Furthermore, the user can choose a time period to study, ranging from January 2004 to present, divided by months or days. The user is also able to compare the RSV of up to five different search terms or the RSV of a particular search term between geographic areas and between time periods. In addition, the user can choose from 25 specific topic categories to restrict the search, each with multiple sub categories for 300 choices in total. With respect to search input, multiple terms could be searched in combination with "+" signs and terms can be excluded with "-" signs. Quotations can be used to specify exact search phrases (Google, 2014). When planning global or local sports events, to attract the most possible attention and interest, organizers must choose appropriate dates. According to the presented methodology and results, GT is a web-based tool that can help.

Conclusion

The described new methodological approach offers bottom-up measurement / comparison for the public interest in certain sports. The interest in a particular sport can also be a valuable indicator for policy makers used to discover sport talent. Is the GT tool and the presented methodology one of the possible solutions to all of the above-mentioned problems closely related to popularity or interest in a particular sport? The results show that football (mean = 30) is the most popular sport according to the requirements of Google trends, as well as the average monthly visit to the sports Wikipedia website in the last 5 years (from November 1, 2015 to October 31, 2020). The other four sports, according to the requirements on GT web browsers, are ranked in the following way - 2. basketball (mean = 8); 3. volleyball (mean = 3); 4. handball (mean = 1); 5. water polo (middle = <1). The results indicate the GT as a new research tool to explore public interests for certain sporting potentials, but with certain limitations. The GT and the methodology presented provide only an arbitrary system of scales, making the comparison qualitative, non-quantitative, limited by the need to pre-specify search terms for a limited comparison in a particular language. Search terms must be as precise as possible, in order to avoid overlaps, and the results will be affected if the difference in the number of queries between the analyzed terms is smaller. A common problem for all data mining research is the existence of digital divisions. Online research tools represent the interests of people who use online resources to pursue their interests in certain sports, while others are excluded. While being aware of the observed shortcomings, we can conclude that GT, as a new research tool, could be a potential solution for conducting research of public interest for a particular sport on a global or regional level, especially in combination with other tools. GT provides practical, useful and sensitive information about the interests of the population for a particular sport. In future research, it will be interesting to explore or predict trends and compare interests between individual and team sports or groups of sports (e.g. water sports and martial arts, winter and summer Olympic sports, etc.). Also, GT and the presented methodology can find their use in sports comparison at national, regional and local level.

Acknowledgments:

This study is part of University of Split, Faculty of Kinesiology project: Development of tools for sport popularity measuring, detecting talents, analysis of training effects and sports achievements,

This study is supported by the Croatian Science Foundation (project number 3366).

References

Baram-Tsabari, A., & Segev, E. (2011). Exploring new web-based tools to identify public interest in science. *Public understanding of science*, *20*(1), 130-143.
Baram-Tsabari, A., & Yarden, A. (2008). Identifying meta-clusters of students' interest in science and their change with age. *Journal of Research in Science Teaching, 46*(9): 999–1022.
Baram-Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2009). Asking scientists: A decade of questions analyzed by age, gender and country. *Science Education, 93*(1): 131–160.
Choi HY, Varian H (2012) Predicting the Present with Google Trends. Economic Record 88: 2–9.
Encyclopedia Britannica, How are SportsChosen for the Olympics<https://www.britannica.com/story/how-are-sports-chosen-for-the-olympics.hr>, accessed 11/15/ 2020.
Falchetti, E., Caravita, S., & Sperduti, A. (2007). What do layperson want to know from scientists? An analysis of a dialogue between scientists and laypersons on the web site Scienzaonline. *Public Understanding of Science,* 16(4): 489-506.
Google (2014) Google Trends Help. https://support.google.com/trends/. Accessed: 10/31/20.
Google (2014) How to type your search term. Available: https://support.google. com/trends/answer/4359582?hl=en&ref_topic=4365599. Accessed 4/25/2014.
Kovačić, K., Petričević, D., & Čular, D. (2017). Sport entrepreneurship, education and secondary career. *Acta Kinesiologica*, *11*(2), 67-71.
Liu, X., & Ruiz, M. E. (2008). Using Data Mining to Predict K–12 Students' Performance on Large-Scale Assessment Items Related to Energy. *Journal of Research in Science Teaching,* 45(5): 554-573.
Mondria, J., Wu, T. (2013) Imperfect financial integration and asymmetric information: competing explanations of the home bias puzzle? Canadian Journal of Economics-Revue Canadienne D Economique 46: 310–337.
Murata, T. (2006, December 18-22). *Towards the Detection of Breaking News from Online Web Search Keywords.* Paper presented at the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Hong Kong.
Murata, T. (2006, December 18-22). *Towards the Detection of Breaking News from Online Web Search Keywords.* Paper presented at the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Hong Kong.
Croatian National SportsProgram (2018)<https://sdus.gov.hr/UserDocsImages/dokumenti/Nacionalni%20program%20%C5%A1porta%202019.%20%E2%80%93%202026_.pdf>, accessed 11/15/2020.
Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of google trends in health care research: a systematic review. *PloS one*, *9*(10), e109583.
Schummer, J. (2005). Reading nano: the public interest in nanotechnology as reflected in purchase patterns of books. *Public Understanding of Science,* 14(2): 163-183.
Segev, E. & Ahituv, N. (2010). Popular searches in Google and Yahoo!: A "digital divide" in the information uses? *The Information Society, 26*(1): 17–37.
Segev, E. & Ahituv, N. (2010). Popular searches in Google and Yahoo!: A "digital divide" in the information uses? *The Information Society, 26*(1): 17–37.
Top End Sports, Sport, Inclusion at the Olympics<https://www.topendsports.com/events/summer/sport-selection.htm,>, accessed 11/15/ 2020.
Topendsports, The World's Most Viewed Sports <*https://www.topendsports.com/world/lists/popular-sport/most-viewed.htm,* accessed 11/15/2020.
Wong, W. K. W. (2012). Faces on Facebook: A study of self-presentation and social support on Facebook.Str. 184-121.

Corresponding information:
Received: 15.11.2020.
Accepted: 03.12.2020.
Correspondence to: Dražen Čular
University: University of Split
Faculty: Faculty of Kinesiology, Split
Phone: +385992509452
E-mail: dcular@kifst.hr