**Introduction of a text mining-based decision support technique**

**Dániel Czúni** (<https://orcid.org/0000-0003-3774-687X>)**,**

Hungarian University of Agricultural and Life Science,

E-mail: [czunidaniel9@gmail.com](mailto:czunidaniel9@gmail.com)

**Gyöngyi Bánkuti**, (<https://orcid.org/0000-0001-9055-1890>)

Hungarian University of Agricultural and Life Science, Institute of Mathematics and Basic Science, Department of Mathematics and Modelling,

E-mail: [bankuti.gyongyi@uni-mate.hu](mailto:bankuti.gyongyi@uni-mate.hu)

**László Pitlik** (<https://orcid.org/0000-0001-5819-0319>)

Kodolanyi Janos University, Institute and Department of Computer Science

E-mail: [pitlik.laszlo@kodolanyi.hu](mailto:pitlik.laszlo@kodolanyi.hu) 

Keywords: FIFA World Cup event Hungary (FIWC), Market analysis for event organization, Similarity Analysis, Component-based Object Comparison for Objectivity (COCO), Excel

Abstract: The aim of this study is to show a text mining-based decision support technique, for in advance decision about the organization of an internationally respected qualification round of PlayStation, namely, FIFA World Cup, (FIWC) in Hungary.

Since the sales data for the game were not available – as that are trade secrets – only the popularity of some relevant search terms related to the FIFA e-football game could be measured to gain information about the market situation in connection to the FIWC event organization. The investigation focused on the 14 countries, that have already delegated competitors to the competitions till the investigated year of 2012, plus Hungary. Fortunately, there were a total of seven international, language-independent search terms, treated with the same spelling in all countries, (“Sony, PlayStation, PlayStation 4, FIFA, FIFA15, EA, EA Sports") for which relevant google trend statistics were available. Only the market situation, none of the other details, like the location of the qualification, financial background, or technical details were considered.

For the analysis, the “similarity analysis” (or also called as COCO, Component-based Object Comparison for Objectivity) technique was used, where the countries were the objects, and the search term expressions were the attributes.

The research concluded the FIFA video game is popular enough to organize the event in Hungary. In the ranking, Hungary became to be the 5th of fifteen countries, means only four countries had better frame conditions to organize the next qualification round.

This paper concludes, the Google Trend Statistic based market analysis – applied for the organization of FIFA World Cup qualification round in Hungary as an example – is an effective decision-making technique.

# Introduction

In our publication, we would like to provide a method for decision preparation in the absence of sales data. To demonstrate this, we rely on previous research, in which the question to be investigated was "Is it worth organizing one of the qualification rounds of the FIFA Virtual Football World Cup, which is part of a well-proven international series of events, in Hungary?" Would there be interest in it, could the event become profitable.

The FIFA series is the most popular e-sports game in the world. Our hypothesis based on our experience was YES, since the game is popular in our country, the aggregate interest in the game is sufficiently high. Popularity means business potential, it primarily attracts endemic companies, sponsors, and fans. Our goal was to investigate the popularity of the game in Hungary. Finding an answer to the question raised during the business planning, according to which event organization would be profitable for Hungary? The 14 countries whose competitors reached the finals in the last three years were included in the investigation. We added Hungary to these. As a result of our research, we want to create a ranking of the popularity of the FIFA game, which will show how relatively popular this game is in our country.

The aim of the investigation was to find out whether it is worth holding an internationally recognized event in Hungary for a qualifying round of the PlayStation game FIFA World Cup. However, game purchase data - from which the actual demand could have been assessed - was not available. Instead, we could rely on Google searches for research, which we think shows how popular a game is in selected countries.

The Object/Attribute Matrix includes as objects the 14 countries whose competitors have reached the finals in the last three years. To these we added Hungary.

As attributes, seven language-independent keyword search terms - Sony, PlayStation, PlayStation 4, FIFA, FIFA15, EA, EA Sports - that were relevant to the topic were selected as attributes, i.e., with the same spelling in all countries, for which data could be successfully extracted for countries from Google Trends.

* Method: Benchmarking + COCO Y0
* Objects: the 15 countries
* Attributes: the 6 search terms in Google Trends.

The data thus defined is placed in the object/attribute matrix. We have used ordinal numbers to indicate how much each term is in demand in that country compared to others.

As a result of the analysis, which was carried out based on the principle that each country is different from the same principle, it can be concluded that Hungary finished in the first third of the set, that is, in 5th place out of 15 countries.

The research proved that there is a sufficiently high aggregate interest in the FIFA game, which proves the popularity of Hungary in the country.

# Literature

The two special branches of data mining are text mining and web mining. Text mining refers to the extraction of knowledge from unstructured or slightly structured text files that was only indirectly hidden in the document file before processing. Data mining is looking for solutions to unstructured problems because we typically look for answers to questions, we can't even ask, which is why we use machine learning. In text mining Unstructured text files are the starting point. The purpose of transforming data is to extract additional data that is valuable to the user.[[1]](#footnote-1)

The two special branches of data mining are text mining and web mining. Text mining refers to the extraction of knowledge from unstructured or slightly structured text files that was only indirectly hidden in the document file before processing. Data mining is looking for solutions to unstructured problems because we typically look for answers to questions, we can't even ask, which is why we use machine learning. In text mining Unstructured text files are the starting point. The purpose of transforming data is to extract additional data that is valuable to the user.[[2]](#footnote-2)

Data quality assurance can be revealed using data mining techniques, which can also be used to support the data cleaning process. [[3]](#footnote-3)

"The expert system is part of artificial intelligence research, it aims to encapsulate human expertise in source code, to use expert knowledge independently of person, space and time. A procedure that allows the consequences of arbitrary factors relationships (state combinations) to be treated in computational form by an intuitive/combinatorial or similarity-based inductive/automatically produced system of relationships. The similarity analysis or decision support system is the expert system, CBR, CRM, abduction, JOKER, COCO, induction, WAM, knowledge-based/knowledge-based system, benchmarking, BSC, fuzzy system, neural net."[[4]](#footnote-4)

The model from which we expect "the most "Good" is the right one. However, this solution does not take into account the social, economic goals of modelling, which may be subjective," but the main aspect to be followed is the general principle of balance creation and preservation."[[5]](#footnote-5)

Online decision support can also be a tool for democratizing the information society since it is used for data and processing methods. Access also gives network citizens the freedom to think (further), (with all the benefits and expected conflicts of this)."[[6]](#footnote-6)

Similarity analysis is a synthesis of the mathematical apparatuses recognized so far, at the same time mathematical statistics (e.g., regression calculus), operations research (e.g., nLP), artificial neural network, decision tree, expert system".[[7]](#footnote-7)

Similarity analysis is a machine version of human case-based inference logics, as well as a possible tool for automating intuition, objective automatic monitoring of balance. In the context of the similarity analysis, quite "bizarre" questions can also be investigated. "The data assets that already exist today and could potentially be further developed can be complemented by appropriate mathematical methods (cf. similarity analysis, case-based inference) and (online) IT solutions that fit workflows." [[8]](#footnote-8)

# Data and methods

To analyse the popularity of the FIFA PlayStation game, similarity analysis - which is a decision support technique based on text mining - and keyword research were applied, since sales data were not available. The frequency of occurrence of relevant search terms related to the FIFA virtual football video game allows us to infer the popularity of the game in the given time interval.

## Similarity analysis (COCO)

The method is suitable for multi-aspect ranking of different alternatives. The aspects are the so-called attributes, and the alternatives are Objects. This is the Object-Attribute Matrix (OAM), the form of which - instead of nominal values - containing the ranks in the columns is used by the COCO Standard Y0 version used for the analysis.

"OAM can be produced by standardization based on primary observation or measurement data. This is the OAM matrix, which contains the evaluation of each object per attribute, already ranked".[[9]](#footnote-9) We also need to specify the value of the direct or inverse proportionality of the attributes and properties, i.e., whether an increase in the characteristic is desirable or a decrease means greater satisfaction - such as in many cases, the price is typical.

The method works with step functions. The parameters are the values of the steps for which a linear programming problem can be written. To solve this, we can use the COCO Y0 method available online or the Excel Solver extension. The model error is smaller with the latter.

To confirm authenticity, a credibility test, a "counter-test", is prepared, which in the normal run is based on an inverse ranking in the opposite direction to the one used. - "is considered credible if the result obtained during the inverse run is contrary to the one in the normal analysis. This is essentially examined by comparing the sign of delta (difference between estimate and fact) values".[[10]](#footnote-10)

The advantages of the COCO method are increasing objectivity, cost-effectiveness, the need for minimal infrastructure, simplicity, wide range of use. The disadvantage of the COCO method in the case of STD is the subjective intervention points - the compilation of the matrix, the selection of the type of run.

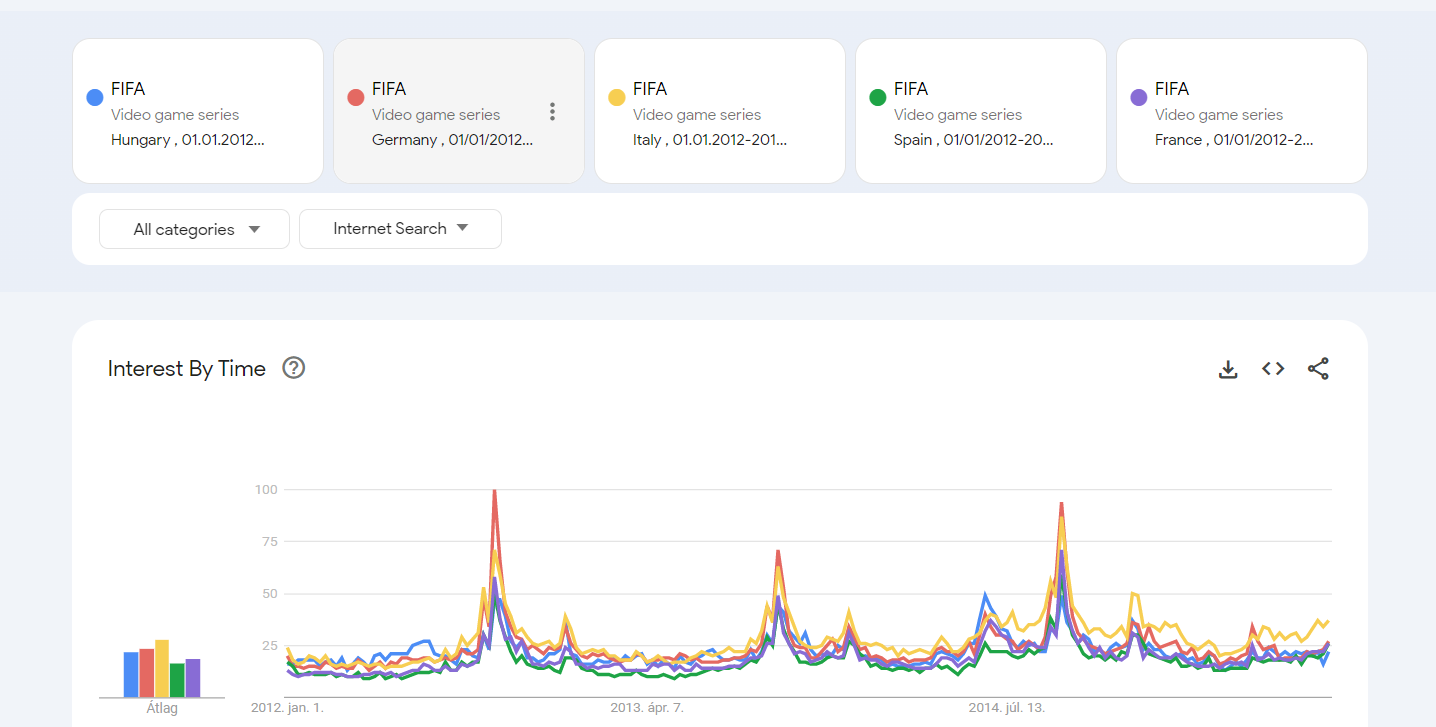
## Google Trends

Google Trends shows you how the number of searches for any search keyword has evolved and gives you statistics on the countries/languages in which the term was searched the most.

"Google Trends advances Google Zeitgeist to open up years of Google search data from around the world, giving you insights from ice cream consumption habits to popularity data from politicians".[[11]](#footnote-11) It is also available in Hungarian. The feature helps you determine the popularity of each search term, the time and geographic origin of your searches, and even what other terms users type them with.

Google Trends gives you the relative values of interest in a search term over the past decade. Relativization returns a ratio of 0-100% relative to itself for a country based on the maximum and minimum interest in the time series. In the case of multiple countries, the ratios are derived from the highest and lowest interest, regardless of the country (c.f. Pitlik, 2022/2023).

Figure#1: Inputs from Google Trends



Source: own presentation (based on the Google Trends URL)

# Discussion

Steps of the analysis:

* Normalization of data Google Trends data (Figure#2)
* Ranking of the percentual raw data (Figure#3)
* Optimizing the anti-discrimination problem in the ranked OAM (Figure#4)
* Comparing naïve and optimized solutions (Figure#5)

Figure#2: Normalized OAM

A képen diagram látható

Automatikusan generált leírás

Source: own presentation – XLS-URL = <https://miau.my-x.hu/miau/297/FIWC%20(10).xlsx>

The zero value above the column name expresses the greater the better orientation as the last parameter of the Excel row number () formula. So, the Figure#2 shows how much the term is searched in which country. To make it easier to find out which keyword is most popular in which country:

* ranking the % interest rates
* colouring the cells by marking the deviations with colours - red is the best, green is the worst one,
* the redder the ordinal number, the more prominent the degree of interest is compared to other objects,
* marking the central country (Germany) in red, comparing other countries to Germany

Figure#3: Ranked OAM

A képen asztal, naptár látható

Automatikusan generált leírás

Source: own presentation – XLS-URL = <https://miau.my-x.hu/miau/297/FIWC%20(10).xlsx>

The so-called naïve (not-optimized) solution obtained by simply averaging the rows of a matrix containing ordinal numbers instead of nominal values in columns arose as a result of row-oriented averaging of the row numbers in Table 2 - Hungary finished in 8th place.

The solution is to copy the ranking table into the data matrix of the Y0 - that is, the anti-discrimination module without row and column headings, by indicating a constant value (1000) in the last column.

Figure#4: Detailed results of the optimized anti-discrimination problem

A képen asztal látható

Automatikusan generált leírás

Source: own presentation - <http://miau.gau.hu/myx-free/coco/test/622132520150703101029.html>,   
*(FIWC (10).xlsx: Sheet/Range: “Eredmény összevetés!A1:L17”)*

*XLS-URL:* [*https://miau.my-x.hu/miau/297/FIWC%20(10).xlsx*](https://miau.my-x.hu/miau/297/FIWC%20(10).xlsx)

As a result of the COCO method, Hungary ranked 5th in the ranking – highlighted in yellow. To compare the naïve (obtained from the simple averaging of the ranking values) order with the COCO result, table 5 is suitable.

Figure#5: Comparison of the naïve and optimized results

A képen asztal látható

Automatikusan generált leírás

Sorce: own presentation   
*(FIWC (10).xlsx: Sheet/Range: “Eredmény összevetés!A1:P17”)*

*XLS-URL:* [*https://miau.my-x.hu/miau/297/FIWC%20(10).xlsx*](https://miau.my-x.hu/miau/297/FIWC%20(10).xlsx)

Based on the final result, Hungary moved up to 5th place after running the objectivity-providing COCO-Y0.

# Conclusion

As a result, we found that Hungary ranked 8th out of the 14+1 countries studied by naïve estimates, while the COCO Y0 method resulted in 5th place.

This means that the supposed popularity of the FIFA game in Hungary was mathematically objectively confirmed based on calculations, that is, it would have been worthwhile to hold the event in our country. The presence of popularity can arouse endemic companies, sponsors. attention, attracting fans, has business value.

It has also been proven that, based on preliminary assumptions, the size of the population of the countries does not necessarily matter to achieve the result, because the size is not necessarily directly proportional to the number of searches for keywords.

Although the FIFA competition event under consideration was not organized for other organizational and financial reasons, the above method can be used in all cases where sales statistics are not available, but relevant keywords can be found that can be used to measure market interest. The purpose of our publication was to present this.

# References - Electronic sources

Adatbányászat *(downloaded: 2015.11.12)*

<http://www.sfact.com/fogalomtar/adatbanyaszat>

Adatbányászat *(downloaded: 2015.11.12)*

<http://www.t-systems.hu/megoldasok/alkalmazasok/analitikus-alkalmazasok-es-uzleti-intelligencia/adatbanyaszat>

Döntéstámogató rendszerek *(downloaded: 2015.11.12)*

<http://ait.gtk.szie.hu/oktatas/pluginfile.php/32452/mod_folder/content/0/7._Dontestamogato_rendszerek.pdf>

Döntéstámogató rendszerek *(downloaded: 2015.11.12)*

<https://prezi.com/z-vqkcrgy-ca/dontestamogato-rendszerek/>

Front-end *(downloaded: 2015.11.12)*

<http://pcforum.hu/szotar/?term=front-end>

Google keresési trendek, immár magyarul is *(downloaded: 2015.11.12)*

<http://www.guglisagok.hu/2009/08/18/google-keresesi-trendek-immar-magyarul-is/>

Google Trends – Wikipedia *(downloaded: 2015.11.12)*

<https://hu.wikipedia.org/wiki/Google_Trends>

Hasonlóságelemzés COCO használatával *(downloaded: 2015.11.12)*

<http://miau.gau.hu/temp/tananyag/ginf/coco_demo.pdf>

Hírek a hasonlóságelemzés világából *(downloaded: 2015.11.12)*

<http://miau.gau.hu/myx-free/>

<http://www.etudasportal.gov.hu/pages/viewpage.action?pageId=18841701>

OECD *(downloaded: 2015.11.12)*

<http://ecopedia.hu/oecd>

Online döntéstámogató rendszerek tervezése *(downloaded: 2015.11.12)*

<http://businessmanagement.hu/index.php?option=com_content&view=article&id=62&Itemid=14&lang=hu>

Pitlik L, - Pető, I.: A konzisztencia-vezérelt jövőmodellek szerepe a stratégiamenedzsmentben *(downloaded: 2015.11.12)*

<http://miau.gau.hu/miau/91/bulletin_hu.doc>

Pitlik L.-Pásztor M. Zs. - Popovics A. - Bunkóczi L. - Pető I.: Online és lokális döntéstámogatási modellek fejlesztési lehetőségei és várható hatásaik *(downloaded: 2015.11.12)*

<https://miau.gau.hu/miau/39/online.rtf>

Pitlik L. (2022/2023): Adatkonszolidáció módszertani demo *(downloaded: 2023.02.12)*

<https://miau.my-x.hu/miau/287/tutorial_google_trends_5plus.xlsx>, <https://miau.my-x.hu/miau/297/google_trends_data_integration_variations.xlsx>

Statisztika *(downloaded: 2015.11.12)*

[www.ksh.hu/statszemle archíve/1999/1999\_05/1999\_05/\_350pdf](file:///D:\Letöltés\www.ksh.hu\statszemle%20archíve\1999\1999_05\1999_05\_350pdf)

Szakértői rendszer - Wikipedia *(downloaded: 2015.11.12)*

<https://miau.gau.hu/mediawiki/index.php/Szak%C3%A9rt%C5%91i_rendszer>

1. http://www.sfact.com/fogalomtar/adatbanyaszat (2015.11.12) [↑](#footnote-ref-1)
2. http://www.sfact.com/fogalomtar/adatbanyaszat (2015.11.12) [↑](#footnote-ref-2)
3. http://www.t-systems.hu/megoldasok/alkalmazasok/analitikus-alkalmazasok-es-uzleti-intelligencia/adatbanyaszat (2015.11.12) [↑](#footnote-ref-3)
4. https://miau.gau.hu/mediawiki/index.php/Szak%C3%A9rt%C5%91i\_rendszer [↑](#footnote-ref-4)
5. http://miau.gau.hu/miau/91/bulletin\_hu.doc (2015.11.12) [↑](#footnote-ref-5)
6. https://miau.gau.hu/miau/39/online.rtf (2015.11.12) [↑](#footnote-ref-6)
7. http://businessmanagement.hu/index.php?option=com\_content&view=article&id=62&Itemid=14&lang=hu (2015.11.12) [↑](#footnote-ref-7)
8. http://businessmanagement.hu/index.php?option=com\_content&view=article&id=62&Itemid=14&lang=hu (2015.11.12) [↑](#footnote-ref-8)
9. Pitlik, et.al. (2008) [↑](#footnote-ref-9)
10. http://miau.gau.hu/temp/tananyag/ginf/coco\_demo.pdf (2015.11.12) [↑](#footnote-ref-10)
11. https://hu.wikipedia.org/wiki/Google\_Trends (2015.11.12) [↑](#footnote-ref-11)