# The measurement errors of google trends data

Google Trends is a popular data source used across various fields, including information systems, business, economics, healthcare, and political science. It examines the popularity of search queries, but its reliability has been questioned due to sampling and measurement errors. While many studies have addressed sampling errors, this article focuses on the measurement errors resulting from changes in Google Trends' data collection methods after 2022.

Google Trends provides search data by normalizing interest levels on a scale from 0 to 100. The highest value represents the peak popularity for a search term during a specific time frame, and other values reflect lower interest in comparison to that peak. However, Google updates its data collection methods regularly, which may introduce new errors, as previous research has pointed out. For example, Lazer et al. were among the first to note potential measurement errors caused by such updates. Since 2022, Google has updated its algorithm to improve accuracy, but this may have also resulted in unexpected errors.

To highlight these measurement errors, this article examines three search terms: ‘covid-19,’ ‘decoupling China,’ and ‘debt trap diplomacy.’ These terms cover topics related to public health, economics, and global politics. The article reveals how Google Trends sometimes produces strange search results for terms that either were not coined yet or provide insufficient data.

**Google Trends Overview**

Google Trends aggregates search interest and normalizes it to allow comparisons across regions and time periods. It shows trends at different temporal resolutions: the last seven days, daily data for the last eight months, weekly data for the last five years, and monthly data for longer periods. The data comes from samples, which means that it does not capture every search but a representative portion of searches.

In some cases, the Google Trends algorithm also uses topic-based searches, which cluster related terms. For example, a search for ‘covid-19’ might include data related to ‘coronavirus’ or ‘Wuhan pneumonia.’ Despite its broad usefulness, researchers need to be cautious of potential errors, especially after Google’s 2022 update.

**Measurement Errors in Google Trends**

To investigate potential errors, three case studies are presented.

**1. Covid-19**

The term ‘covid-19’ was officially coined by the World Health Organization (WHO) in February 2020. However, Google Trends shows search results for this term before 2020. Searches for ‘covid-19’ appear as far back as 2015, which is implausible since the virus did not exist then. Some scholars suggest that this may be due to sampling errors, while others argue that Google Trends combines older content with new data, thus misleading the algorithm.

Despite using many samples to correct errors, the search results still show data for the term ‘covid-19’ before it was known. This implies that the issue is not solely related to sampling but also to Google’s data collection methods.

**2. Decoupling China**

The term ‘decoupling China’ refers to the efforts by countries, particularly the US, to reduce economic dependence on China. Search interest for this phrase spiked around 2018, when the US-China trade war began. However, Google Trends shows search data for ‘decoupling China’ before 2018, which is unlikely, as the term was not widely used before the trade tensions between the two countries escalated.

Similar to the case of ‘covid-19,’ these results suggest that Google Trends may be erroneously linking older searches with newer content, which introduces misleading data. Even though Google Trends updated its algorithms, these errors persist.

**3. Debt Trap Diplomacy**

‘Debt trap diplomacy’ refers to the idea that a country intentionally extends excessive credit to another country to gain political leverage. This phrase has been commonly used in discussions about China’s Belt and Road Initiative, particularly after 2017. However, as in the previous cases, Google Trends shows search interest for ‘debt trap diplomacy’ before the term gained prominence.

These examples illustrate the presence of measurement errors in Google Trends data after 2022. It appears that the update intended to improve the accuracy of search results may have inadvertently caused errors by linking new and old data inappropriately.

**Implications for Researchers**

Measurement errors in Google Trends data have significant implications for research, particularly in fields like economics, healthcare, and international relations. Many studies rely on Google Trends to forecast trends, predict behavior, and understand public interest in various issues. If the data are inaccurate or misleading, it could result in flawed research findings.

For instance, using incorrect search data for ‘covid-19’ might distort studies on the public’s awareness and response to the pandemic. Similarly, measurement errors in searches related to US-China relations could affect research on economic policies or geopolitical strategies. Scholars must be careful when using Google Trends data after 2022 and take steps to validate the results through other means, such as replicating studies conducted before the update.

In particular, researchers in fields like business and healthcare should consider re-running analyses using data from both before and after 2022 to assess the impact of these errors. The article suggests that researchers take these steps to ensure the accuracy of their findings and avoid making erroneous conclusions based on flawed Google Trends data.

**Conclusion**

Google Trends remains a valuable tool for analyzing public interest in various topics. However, changes in its data collection methods since 2022 have introduced potential measurement errors that scholars must be aware of. This article demonstrates that search results for terms like ‘covid-19,’ ‘decoupling China,’ and ‘debt trap diplomacy’ show anomalies that raise concerns about the data’s accuracy.

Scholars are advised to replicate their research and validate the Google Trends data they use, especially for forecasting in critical fields like economics and healthcare. Further investigation is needed to fully understand the scope of these measurement errors and how they might impact various fields of study. Researchers must continue to exercise caution and seek alternative data sources to corroborate their findings.

**Dictionary**

1. **Sampling errors** – Ошибки выборки
2. **Measurement errors** – Ошибки измерения
3. **Temporal resolution** – Временное разрешение
4. **Algorithm** – Алгоритм
5. **Normalize** – Нормализовать
6. **Cluster** – Объединять в группы, кластеризовать
7. **Improbable** – Маловероятный, неправдоподобный
8. **Anomalies** – Аномалии
9. **Geopolitical strategies** – Геополитические стратегии
10. **Corroborate** – Подтверждать, подкреплять (данные или гипотезу)

### 15 вопросов:

1. What fields commonly use Google Trends data?
2. What types of errors does this article focus on regarding Google Trends data?
3. How does Google Trends normalize search interest data?
4. When did Google update its data collection methods, according to the article?
5. What are the three example search terms analyzed in the article?
6. What is the anomaly found in Google Trends search results for ‘covid-19’?
7. What explanation is suggested for the false ‘covid-19’ search results before 2019?
8. What were the peaks of search interest for ‘decoupling China’ attributed to?
9. How did the US-China trade war influence search trends for ‘decoupling China’?
10. What is the main conclusion drawn about measurement errors in Google Trends?
11. How do these measurement errors affect research in fields like economics and healthcare?
12. What potential consequences could these errors have for policymakers or businesses?
13. What does the article suggest scholars do when using Google Trends data after 2022?
14. How do the search anomalies for ‘decoupling China’ relate to the concept of sampling errors?
15. How does the article suggest Google Trends data may affect forecasting accuracy?

Реферат:

Статья рассматривает проблемы измерения данных Google Trends, особенно после изменения метода сбора данных в начале 2022 года. Google Trends, популярный инструмент для анализа запросов, может иметь ошибки, влияющие на достоверность получаемых результатов.

В исследовании выделяются три ключевых термина: «covid-19», «decoupling China» и «debt trap diplomacy». Например, результаты по запросу «covid-19» показывают интерес к этому термину задолго до его официального появления в 2020 году. Это свидетельствует о наличии систематических ошибок, которые не устраняются даже при использовании больших выборок. В случае термина «decoupling China» поисковые результаты показывают активность пользователей до начала торговой войны в 2018 году, что также вызывает вопросы о точности данных.

Эти аномалии могут искажать выводы исследователей в области международных отношений и экономики. Статья подчеркивает, что подобные ошибки могут привести к неверным прогнозам и плохим решениям в таких областях, как бизнес и здравоохранение.

Авторы призывают ученых повторно проверять свои выводы, используя данные как до, так и после 2022 года, чтобы лучше понять влияние этих изменений. В заключение, статья акцентирует внимание на необходимости тщательной проверки данных Google Trends, а также на возможности использования альтернативных источников для повышения надежности исследований.

Эта работа служит предупреждением для исследователей о том, как важно быть внимательными при использовании таких инструментов в научной деятельности.

Пересказ:

The article explores the measurement errors found in Google Trends data, especially following the 2022 update to its data collection methods. Google Trends is widely used for analyzing search interests, yet its reliability can be compromised by various errors. It focuses on three important terms: ‘covid-19,’ ‘decoupling China,’ and ‘debt trap diplomacy.’ The data for ‘covid-19’ reveals interest in the term from years before its official naming in 2020, indicating systematic issues with the data. Similarly, searches for ‘decoupling China’ show user interest prior to the onset of the US-China trade war in 2018, raising concerns about the data’s accuracy. Such anomalies can lead researchers astray, especially in fields like economics and international relations. The authors argue that these measurement errors might lead to incorrect forecasts and decisions in sectors such as business and healthcare. Researchers are encouraged to reassess their studies using data from both before and after 2022 to gauge the impact of these discrepancies. The article concludes with a call for careful validation of Google Trends data and suggests the exploration of alternative data sources to enhance research reliability. This work serves as a cautionary tale for scholars regarding the critical importance of diligence when utilizing these analytical tools in their studies.