

**Эконометрический и кластерный анализ зарубежных вузов с использованием данных Google Trends.**

**Научный руководитель – Стефановский Дмитрий Владимирович**

***Ахrameев Максим Дмитриевич***

*Студент (магистр)*

Российская академия народного хозяйства и государственной службы при Президенте  
Российской Федерации, Экономический факультет, Москва, Россия

*E-mail: akhrameevmd@gmail.com*

Активность пользователей в интернете уже давно анализируется, и для этого созданы различные сервисы и инструменты. Одним из таких сервисов является Google Trends, который предоставляет данные об относительных характеристиках поисковых запросов в динамике. С использованием данного сервиса были получены недельные временные ряды поисковых запросов о крупнейших зарубежных вузах, с целью выявления трендов и анализа изменения популярности, а также для выделения сезонных компонент.

Google Trends одновременно предоставляет только относительные данные для 5 разных запросов, поэтому встает проблема сравнения в случае количества запросов более пяти. Для формирования общего рейтинга необходимо попарно сравнить запрос с каким-либо одним базовым запросом. В исследовании базовым запросом был выбран «Caltech». В результате был сформирован общий рейтинг популярности, в частности, в топ-3 вузов по количеству поисковых запросов вошли MIT, Harvard, Oxford, причем MIT побеждает с огромным отрывом (более чем в 10 раз), а Oxford в 2 раза популярнее Harvard.

После получения рейтинга, к каждому временному ряду запросов был применен метод TBATS[1], основанный на экспоненциальном сглаживании, для выделения трендов и сезонной составляющей. Также была проанализирована структура этих данных с использованием методов кластеризации. Был проведен анализ качества кластеризации для количества кластеров, начиная от 2 до 10, и наилучшее качество было показано при трёх кластерах. Стоит отметить тот факт, что вузы из первой тройки полученного рейтинга попали в разные группы. Это свидетельствует о том, что администрации данных вузов проводят приемные кампании в разные периоды, что может означать попытку уйти от конкуренции за студентов и позволяя студенту участвовать в разных вступительных испытаниях.

Данный подход обработки поисковых запросов может быть использован в различных областях, а инструменты, предложенные в исследовании, позволяют организовать автоматизированную обработку большого количества данных.

**Источники и литература**

- 1) Alysha M De Livera, Rob J Hyndman and Ralph D Snyder. // Forecasting time series with complex seasonal patterns using exponential smoothing. // Journal of the American Statistical Association 106(496), pp 1513-1527