

Мальцева Е.И.

*Научный руководитель - к.т.н., доц. каф ИСПИ Г.Е. Монахова
Владимирский государственный университет имени А.Г. и Н.Г. Столетовых (ВлГУ)
600000, г. Владимир, ул. Горького, 87
kat.malceva@yandex.ru*

Обзор методов визуализации данных социальных сетей

В современном мире нас окружает огромное количество информации, с чем связано повышение интереса к такому вопросу, как визуализация данных.

Такой резкий скачок к потреблению визуальной информации связан с тем, что сознание не способно успеть воспринять, отсортировать и проанализировать весь поток информации [1]. Визуально представленная информация более привычна и понятна человеческому глазу, с помощью нее можно быстро донести любые мысли и идеи.

Социальные сети представляют собой огромный набор разнородных данных, и, соответственно, широкие возможности для визуализации. В октябре 2020 года число активных авторов в социальных медиа в России составило 64 млн. Авторы написали более 1,2 млрд публичных сообщений (постов, репостов и комментариев) [2].

Существуют различные способы получения данных из социальных сетей, и основным из них является использование API. С их помощью можно получать данные из таких популярных социальных сетей, как ВКонтакте, Instagram, Twitter и т.д.

С ростом востребованности области визуализации данных появляется необходимость в ориентировании среди многообразия доступных методов [3]. Выбор метода визуализации не так прост. Одну и ту же информацию можно представить по-разному, используя те или иные средства. Однако для того, чтобы визуализация выполняла свою функцию представления информации в простом и доступном для восприятия виде, необходимо придерживаться различных правил. Например, необходимо сделать так, чтобы при взгляде на визуальное представление информации можно было сразу выявить закономерности в исходных данных и принимать на их основе решения.

Экономист, специалист по визуализации данных и презентациям Джон Швебиш сформулировал четыре основных принципа визуализации [4]:

- 1) Ясность данных: не стоит перегружать схемы лишней информацией — лучше пусть ее будет меньше, но данные будут проверенными и понятными;
- 2) Меньше визуального шума: необходимо избегать темных или тяжелых линий сетки, лишних иконок и меток, большого количества текста, теней и градиентов, чрезмерного объема;
- 3) График и текст — единое целое: необходимо, чтобы графики и текст дополняли друг друга;
- 4) Характеристики, которые можно определить сразу: некоторые визуальные характеристики, такие как форма и контраст, мы можем заметить и определить сразу, не особо всматриваясь в изображение, поэтому с их помощью стоит выделять какие-либо важные данные на фоне других.

Выбор того или иного метода визуализации во многом зависит от поставленной задачи и от характера набора данных.

В качестве наборов данных можно выделить следующие:

- массив данных;
- социальный граф (хранит информацию об узлах и соединяющих их ребрах);
- карта (отличительным признаком является наличие географических координат).

Говоря о задачах визуализации, то есть о том, с какой целью данные визуализируются, обычно выделяют следующие цели [5]:

- сравнить показатели;
- показать распределение данных;
- показать структуру (состав);
- проследить взаимосвязь между показателями.

Рассмотрим несколько примеров. Среди двухмерных средств для визуализации массивов данных наиболее широко известны:

1) Линейный график - помогает отобразить тенденцию, передать изменения какого-либо признака во времени (графики строятся по осям X и Y и показывают зависимость данных друг от друга, при этом для сравнения нескольких рядов чисел графики наносятся на одни и те же оси координат);

2) Гистограммы - применяются для сравнения значений в течение некоторого периода или же соотношения величин;

3) Горизонтальные столбиковые диаграммы - обычно используются, когда нужно сравнить большое количество показателей или визуально выделить явное превосходство одного из них;

4) Круговые диаграммы - используют, если необходимо отобразить соотношение частей и целого, т.е. для анализа состава или структуры явлений;

5) Диаграммы рассеяния - используются для выявления корреляции между двумя переменными;

6) Пузырьковые диаграммы - используются для сравнения и отображения взаимосвязей между разными переменными с помощью их местоположения и пропорций (часто используется для анализа паттернов и поиска корреляций);

7) Тепловые карты — графическое представление данных, где индивидуальные значения в таблице отображаются при помощи цвета, что можно использовать как для определения взаимосвязей между данными, так и для демонстрации изменений данных за определенный период времени (если один из рядов или столбцов содержит временные интервалы);

8) Поточковые графики - отображают изменения данных разных категорий во времени с помощью плавных, взаимосвязанных форм, которые чем-то напоминают течение реки, благодаря чему эстетичны и более приятны для восприятия.

Для визуализации связей между пользователями социальной сети чаще всего используют социальные графы. Социальный граф — граф, узлы которого представлены социальными объектами, такими как пользовательские профили с различными атрибутами (например, имя, день рождения, родной город), сообщества и так далее, а рёбра — социальными связями между ними. Также можно использовать диаграмму Венна – это диаграмма, которая визуально отображает все возможные логические отношения множеств, каждое из которых, как правило представлено окружностью. Другие примеры - хордовая и неленточная хордовая диаграммы.

При визуализации геолокационных данных обычно используют карты, такие как пузырьковая карта (удобна для сравнения пропорциональных значений в разных географических регионах), фоновая картограмма (используется для разделения по географическим областям или регионам, выделенным разным цветом в соответствии с переменными данными), карта взаимосвязей (используются для отображения взаимосвязей и взаимоотношений в пределах географической территории), точечная карта (показывает распределение данных по географическим регионам).

Таким образом, методов визуализации существует очень много и выбор конкретного метода зависит от решаемой задачи и имеющихся в распоряжении данных.

Литература

1) Борисов, Е. Е. Визуализация как актуальное направление распространения информации / Е. Е. Борисов. — Текст : непосредственный // Молодой ученый. — 2019. — № 22 (260). — С. 611-614. — URL: <https://moluch.ru/archive/260/59960/>

2) Черный, В. Социальные сети в России: цифры и тренды, осень 2020 / В. Черный — URL: <https://br-analytics.ru/blog/social-media-russia-2020/>

3) Северино, Р. Каталог визуализации данных / Р. Северино — URL: <https://datavizcatalogue.com/RU/index.html>

4) Визуализация данных: способы, инструменты, полезные ссылки — URL: https://blog.adn.agency/article/vizualizatsiya_dannykh_sposoby_instrumenty_poleznye_ssytki

5) Иванова, А., Чаусова, А., Миргородская, О. Визуализация данных: основные правила, полезные приемы и инструменты / А. Иванова, А. Чаусова, О. Миргородская — URL: <https://www.owox.ru/blog/articles/data-visualization/>