

УДК 004.5

Подбор релевантной информации для группы пользователей на основе анализа их интересов

Автор: Морозова П. И., Университет ИТМО, Санкт-Петербург.

Научный руководитель: Тесля Н.Н., к.т.н., Университет ИТМО, Санкт-Петербург

Цифровые вывески используются повсеместно для демонстрации рекламы и новостей, обмена информацией между членами сообщества. Цифровые вывески можно найти в различных учреждениях, офисных зданиях, торговых центрах.

Цели и местоположения вывесок определяют требования и ограничения, на основе которых строятся алгоритмы их работы. В данной работе будут рассматриваться цифровые вывески, установленные в торговых центрах. Цель установки подобных вывесок – демонстрация информации, в том числе рекламного характера, зрителям, находящимся перед вывеской. Поскольку системы управления вывесками могут выбирать информацию для демонстрации, существует возможность предоставления информации в зависимости от того, кому он предназначена, т.е. предоставления персонализированной информации. Персонализация демонстрируемой таким образом информации может существенно повысить заинтересованность зрителей, и, следовательно, эффективность работы вывески.

В данной работе будет проведен анализ существующих подходов к проектированию подобных вывесок, а также будет предложена концепция системы предоставления персонализированной информации посредством цифровых вывесок, призванная улучшить существующие подходы.

При проектировании цифровых вывесок необходимо учитывать ограничения, накладываемые законами и регламенты о работе с персональными данными. Например, 152 ФЗ на территории Российской Федерации или GDPR на территории Евросоюза. Эти законы, в частности, налагают запрет на сбор, хранение и обработку персональных данных пользователя без его письменного согласия.

Поскольку цифровые вывески, рассматриваемые в данной работе, установлены в торговых центрах, собрать согласия посетителей не представляется возможным. Для решения данной проблемы предлагается использование анонимной информации о зрителях. В классических работах к использованию предлагаются следующие характеристики: пол и возраст. При этом, как правило, зрители делятся на шесть демографических групп: юные, взрослые и пожилые мужчины и женщины.

Мы предлагаем, во-первых, разбивать зрителей на группы по возрасту таким образом, чтобы разница в возрасте между двумя членами группы не превышала 10 лет. Во-вторых, мы предлагаем учитывать эмоции пользователя. Современные технологии позволяют определять вероятности того, что зритель находится в одном из восьми эмоциональных состояний. Данный подход позволит увеличить количество демографических групп и более точно подбирать темы, интересные той или иной группе.

Кроме того, возникает вопрос о том, как предоставить информацию, которая была бы интересна представителям нескольких демографических групп. Существующие подходы предлагают выбрать одну группу и продемонстрировать информацию, интересную представителям именно этой группы пользователей.

Мы в свою очередь предлагаем разбить подбор релевантной информации на несколько этапов:

1. Определение зрителей, находящихся в данный момент перед цифровой вывеской, и отнесение их к соответствующим демографическим группам.
2. Оценка интереса представителей определенной демографической группы к той или иной теме. Для оценки данного показателя предлагается найти усредненное процентное значение времени, которое представители данной демографической группы как правило тратят на просмотр информации, относящейся к данной теме.
3. Вычисление значения интереса для темы, информация о которой ранее не была показана представителям данной демографической группы, на основе определенных интересов представителей данной демографической группы. Для этого предлагается использовать онтологию, представленную в виде взвешенного ориентированного графа. Предлагается задать следующие ограничения на веса для дуг графа:
 - Дуга от родительского концепта к дочернему концепту имеет вес 0.
 - Остальные дуги имеют вес от 0 до 1.Затем определяется расстояние между каждой парой концептов. На основе этих расчётов вычисляется значение интереса для неизвестных данной демографической группе тем.
4. Расчет заинтересованности всех присутствующих в информации по каждой теме. На основе результатов данного расчета подбирается отображаемая информация.

Для того, чтобы убедиться в возможности применения данной концепции, было проведено исследование, в рамках которого опрашиваемым демонстрировались короткие анимации, которые должны были вызвать ту или иную эмоцию.

В качестве программного обеспечения использовался Microsoft Azure Face API. Было показано, что данное программное обеспечение позволяет определить пол со 100% точностью и возраст с точностью +/- 4 года. Также было показано, что в большинстве случаев удалось повлиять на эмоциональное состояние людей с помощью демонстрации соответствующих анимаций.

В работе предложена новая концепция работы систем управления рекламными вывесками в торговых центрах, проведен эксперимент, показывающий, что современные технические средства предоставляют достаточную точность для реализации данной концепции.