

# САМООРГАНИЗУЮЩИЙСЯ ВЕБ-САЙТ

*Self-organizing website*

**Аксенов А. В.**

*Санкт-Петербургский Государственный Университет Аэрокосмического  
Приборостроения*

*автор – студент кафедры «Вычислительных систем и сетей»*

*науч.рук. – ст.преп. Козловский А. Ю.*

## Аннотация

Рассматривается принцип построения сайта, анализирующего деятельность пользователей для автоматического определения ценности контента.

Количество информации в Интернете растет с каждой секундой, чего нельзя сказать о ее качестве. Именно поэтому все более остро стоит проблема «очистки зерен от плевел», т. е. отбора информации (контента), действительно несущей ценность и заслуживающей внимания. Для этого существует несколько способов.

Можно отбирать контент, анализируя его содержание. Такой принцип лежит в основе работы спам-фильтров и эффективен далеко не всегда. Иной способ предполагает использование модераторов – пользователей, в чьи обязанности входит контроль контента на предмет соответствия его некоторым оговоренным для конкретного сайта правилам, общепринятым нормам морали и этики, а также личным взглядам самого модератора. Такой подход является плохо масштабируемым и зачастую (при больших объемах информации) нецелесообразным.

В Интернете существует множество так называемых «управляемых пользовательским сообществом» или «демократических» веб-сайтов. Их основной принцип «коллективного определения ценности информации» заключается в том, что контент отбирается самими пользователями, а не специально выделенными модераторами, путем введения рейтингов пользователей и их сообщений. Но такой подход также не лишен серьезных проблем.

Большинство этих сайтов страдают от того, что происходит выделение «элиты», т. е. пользователей с чрезвычайно высоким рейтингом (как правило, их число составляет менее 1 % общего количества пользователей), которые влияют на жизнь сайта, при этом сводя результаты деятельности большинства к нулю [1]. Эта проблема обусловлена тем, что каждый пользователь может напрямую влиять на рейтинг другого прямым голосованием, что позволяет пользователям объединяться в группы для достижения своих, как правило отрицательных для сайта и других пользователей, целей.

В основе данной работы лежит отказ от прямой манипуляции рейтингом в пользу расчета рангов пользователей и информации самим сайтом на основе анализа их авторской деятельности, активности в обсуждениях, ссылок на сообщения других авторов.

Для ранжирования пользователей и их сообщений выбран аналог алгоритма PageRank, разработанного основателем компании Google Л. Пейджем [2] для анализа информационной ценности веб-страниц. Его суть состоит в анализе направленного графа, вершинами которого являются страницы, а ребрами – гиперссылки между ними с получением в качестве выходных данных весов страниц, влияющих на порядок вывода результатов запроса к поисковой системе компании. В данной работе модифицированный алгоритм используется для решения иных задач, работая с графом, вершинами которого являются пользователи сайта и их сообщения, а ребрами – отношение, которые между ними возникают, в частности между автором и его сообщением, между сообщением и ответом на это сообщение, между автором удаленного сообщения и пользователем, его удалившим, и другие.

## Библиографический список

1. Wilson C., *The Wisdom of the Chaperones: Digg, Wikipedia, and the myth of Web 2.0 democracy.* <http://www.slate.com/id/2184487/>
2. Page L., *The PageRank Citation Ranking: Bringing Order to the Web* <http://infolab.stanford.edu/~backrub/pageranksub.ps>