
Исследование экологических сайтов

Луговая Н. Б., Печников А. А.

ИПМИ КарНЦ РАН, ул.Пушкинская 11, Петрозаводск, 185910, Россия

e-mail: nataly@krc.karelia.ru, pechnikov@krc.karelia.ru

Web-сайты экологической направленности играют все большее значение в российском фрагменте Интернета. В Яндекс.Каталоге [1], где наиболее значимые сайты по разделам отбираются вручную, в настоящее время представлено более 130 сайтов по разделу «Экология». Представляет большой научный и практический интерес вопрос о том, насколько связаны сайты экологической направленности между собой.

Назовем коммуникационным графом сообщества сайтов (или коротко web-графом) ориентированный взвешенный мультиграф $G=(S, L, t, w)$, где S – множество вершин, соответствующих сайтам сообщества, L – множество дуг, соответствующих гиперссылкам сайтов сообщества друг на друга, t и w – весовые функции, определенные на множествах S и L соответственно. Формальный анализ web-графа позволяет дать ответы на ряд содержательных вопросов, таких, например, как степень связности участников сообщества.

В качестве множества S (целевой группы исследования) взяты 8 экологических сайтов, имеющих наибольший тематический индекс цитирования (ТИЦ) в поисковой системе Яндекс: «Всемирный фонд дикой природы в России», «Социально-экологический союз», «Леса России», «Центр охраны дикой природы», «Всероссийский экологический портал», «Байкальская экологическая волна», «Природа России» и «Biodat - сохраним природу вместе».

Процедура построения web-графа сообщества экологических сайтов заключается в следующем. С помощью программы-краулера, являющегося оригинальной разработкой авторов, были отсканированы все гиперссылки, исходящие с сайтов множества S , выявлены ссылки друг на друга и построено множество L . Весовая функция t определена как $t(lp)$, где $lp=0,1,2,\dots$ – уровень страницы, на которой находится соответствующая ссылка; $t(lp1) > t(lp2)$, если $lp1 > lp2$. В качестве весов вершин t были приняты измеренные значения ТИЦ для каждого сайта.

Анализ построенного web-графа позволяет сделать ряд содержательных выводов относительно взаимодействия сайтов экологической направленности. Например, наблюдается очень высокая степень связности web-графа, однако при этом имеется сайт, на который отсутствуют ссылки с других участников целевой группы, и сайт, не имеющий ссылок на других участников.

Литература

Яндекс. Каталог. [Электронный ресурс] – 2008. – Режим доступа: <http://yaca.yandex.ru/yca/cat/>.