

М. М. Соловьев, И. А. Санников (Ульяновск, УлГУ). Система сбора данных в случайной среде для оценивания параметров поисковых структур.

В работе, представленной данным сообщением, описывается система сбора данных для оценивания параметров поисковых структур. В качестве поисковой структуры выбран Интернет-поисковик «Яндекс». Полученные данные сохраняются и обрабатываются в реальном времени. Параллельно строятся траектории изменения параметров во времени и по мере накопления наблюдений уточняется оценка параметров. Сложность задачи состоит как в сборе данных (из-за возможной недоступности поисковой структуры или объекта поиска), так и в анализе полученных данных (их неполноты, дискретности, неточности). Но несмотря на это, можно определить, в связи с чем меняется значимость ресурса в поисковой структуре при изменении параметров объекта. Дальнейший анализ позволяет формализовать поисковую структуру, выявить и оценить значимые параметры объекта.

Формально задачу можно представить следующим образом. Пусть $\{p_i\}$ — множество параметров, используемых поисковой структурой ($i \in [1, n]$, где n — число параметров). Например, p_i — количество внешних ссылок на странице или значение тематического индекса цитирования в Яндексе, $\{k_i\}$ — множество коэффициентов (так называемого веса) для этих параметров. Тогда по результатам $\{I^t\}$ — множество значимостей объекта поиска со временем t ($t \in [1, m]$, где m — число наблюдений, $m \geq n$ и $m \leq 100$), а I^t , например, позиция ресурса при поисковом запросе. Далее строится линейное приближение (1):

$$I^1 = \sum_{i=1}^n p_i k_i, \quad i \in [1, n], \quad \dots, \quad I^m = \sum_{i=1}^n p_i k_i, \quad i \in [1, n], \quad (1)$$

и определяются коэффициенты $\{k_i\}$.

Стоит отметить, что значение веса того или иного параметра определяется лишь с некоторой ошибкой. Чтобы добиться лучшей точности оценки при построении линейного, квадратичного, экспоненциального и некоторых иных приближений, необходимо увеличить количество результатов $\{I^t\}$ рассматриваемых объектов в рамках одной поисковой структуры при различных запросах.

Для проверки полученных результатов в поисковую структуру внедряется новый объект с заданными параметрами, для которого вычисляется теоретическая значимость, после чего сравнивается с реальной.

Программный комплекс выполнен в виде сайта на языке программирования PHP. Интерфейс сайта легко позволяет получить необходимую информацию, доступ к которой осуществляется после ввода имени пользователя и пароля администратора.

Автор выражает благодарность за внимание к работе профессору А. А. Бутову.