

Сервис вебометрического анализа структур научных сайтов

В. В. Наумова¹ и К. А. Платонов¹

Получено 28 февраля 2013 г.; опубликовано 8 марта 2013 г.

В статье описано создание интернет-сервиса, осуществляющего мониторинг научных ресурсов Сети, включающего в себя механизмы автоматической интеграции сведений о научных сайтах, а также сервисы их обработки. Сбор данных осуществляется в автоматическом режиме с использованием API известных поисковых систем (Google, Bing, Yahoo и Яндекс). При ранжировании сайтов вычисляется “вебометрический ранг” сайта. Для анализа структур научных сайтов использованы математические методы. **КЛЮЧЕВЫЕ СЛОВА:** Научный сервис; вебометрика; научный сайт; Российская академия наук.

Ссылка: Наумова, В. В. и К. А. Платонов (2013), Сервис вебометрического анализа структур научных сайтов, *Вестник ОНЗ РАН*, 5, NZ3001, doi:10.2205/2013NZ000117.

Применение вебометрики для анализа научных интернет-ресурсов

В 1997 году Tomas C. Almind и Peter Ingwersen в работе “Informetric analyses on the World Wide Web: Methodological approaches to webometrics” [Almind and Ingwersen, 1997] ввели термин “webometrics”, обозначающий раздел информатики, в котором исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий применительно к World Wide Web.

Доступ к научной информации глобальной сети Интернет для пользователей осуществляется, в основном, с помощью поисковых систем. Наличие высоких рейтинговых оценок для сайта крайне важно с точки зрения его более высокой доступности для широкой аудитории пользователей сети Интернет, поскольку рейтинговые оценки используют практически все поисковые системы. Наличие рейтинговых систем для сайтов, разрабатываемых и применяемых крупными поисковиками (Google, Яндекс и др.), при выдаче результатов поиска позволяет пользователям всего мира быстрее находить наиболее качественную и отвечающую запросу информацию. Результаты запросов пользователей сортируются поисковыми машинами и предоставляются пользователям в порядке уменьшения их рейтинговых оценок. Таким образом, на-

личие высоких оценок ставит сайт в лучшее положение по отношению к другим сайтам.

Самый известный проект в области вебометрики реализовала группа Кибернетической лаборатории Центра научной информации и документации Национального исследовательского совета Испании “Webometrics Ranking of World Universities” (Ranking Web of World Research Centers, <http://research.webometrics.info>).

Используя методы вебометрики, лаборатория каждые полгода публикует на своем сайте и мировой рейтинг сайтов научно-исследовательских организаций (Ranking Web of World Universities, <http://www.webometrics.info>).

Российские работы в данном направлении проводились Институтом вычислительных технологий Сибирского отделения РАН и Институтом прикладных математических исследований КарНЦ РАН. За основу были приняты методы и подходы Кибернетической лаборатории Испании, но с некоторыми изменениями и дополнениями. Исследователями Сибирского отделения РАН создан ряд интернет-сервисов для вычисления рейтингов научных сайтов Сибирского отделения РАН, которые включают: выбор формулы, разбиение по направлениям исследований, выбор группы исследуемых сайтов [Шокин и др., 2008].

Работы Института прикладных математических исследований КарНЦ РАН выполнены с использованием теории графов, которая применялась для анализа структуры интернет-пространства научных организаций РАН северо-запада России, а также других подмножеств научных и университетских сайтов России. На основе результатов авторами построена модель “научного веба” [Маза-лов и Печников, 2009; Печников, 2008, 2010].

Предложенная модель позволяет увидеть администра-

¹Дальневосточный геологический институт ДВО РАН, Владивосток, Россия

тивный каркас, тематические связи, выделить внутренние ресурсы сайта. Проведенные исследования позволили выявить особое множество веб-коммуникаторов: сайты-индукторы (“мало входящих ссылок, много исходящих”), сайты-коллекторы (“много входящих, мало исходящих”), сайты-посредники (“много входящих, много исходящих”).

В Дальневосточном геологическом институте ДВО РАН работы по анализу интернет-пространств научных сайтов проводятся с 2009 года [Наумова, 2012; Ханчук и Наумова, 2009], при этом базовым множеством анализа были выбраны сайты институтов Дальневосточного отделения РАН.

Интернет-ресурсы научных институтов РАН представляют собой сложный информационный комплекс. Как правило, в этом комплексе имеется официальный сайт института, а также независимые сайты лабораторий, научных журналов, электронных библиотек, страницы научных сотрудников, информационные системы, ГИС-порталы, сайты конференций, семинаров, тематические ресурсы и т. д.

Логичным продолжением работ, ведущихся в этом направлении в лаборатории информационных технологий Дальневосточного геологического института ДВО РАН, стало создание интернет-сервиса <http://webometrics.fegi.ru>, осуществляющего мониторинг научных ресурсов Сети и включающего в себя механизмы автоматической интеграции сведений о научных сайтах, а также сервисов их обработки.

Сбор данных

Сбор данных осуществляется в автоматическом режиме с использованием API известных поисковых систем (Google, Bing, Yahoo и Яндекс). Данные включают:

1. количество уникальных гипертекстовых ссылок с других ресурсов (V – visibility, цитируемость);
2. общее количество страниц сайта (S – size, размер);
3. количество полнотекстовых файлов, под которыми понимаются файлы с расширениями pdf, ps, doc, xls, ppt и rtf (R – “rich files”);
4. научность сайта (Sc – “shoolar”).

Ранжирование и анализ структур научных сайтов

При ранжировании сайтов рассчитывается ранг объекта по каждому индикатору. ($RankS$, $RankV$, $RankR$, $RankSc$), далее вычисляется интегральный показатель, называемый “вебометрический ранг” (WR – Webometrics

Rank), который получается в результате ранжирования суммы рангов, умноженных на коэффициенты [Aguillo et al., 2006]:

Webometrics Rank (position) =

$$4 \times RankV + 2 \times RankS + 1 \times RankR + 1 \times RankSc$$

Для анализа структур научных сайтов использованы методы теории графов и метод главных компонент.

В настоящее время на сервисе представлены результаты мониторинга информационных пространств Отделения наук о Земле РАН и Дальневосточного отделения РАН.

Анализ информационного пространства Отделения наук о Земле РАН

Множество Отделения наук о Земле РАН включает в себя 67 сайтов институтов (<http://webometrics.fegi.ru/index.php?Itemid=3>). На сервисе представлены результаты мирового рейтинга по состоянию на 7 февраля 2013 г. (http://webometrics.fegi.ru/index.php?option=com_content&view=article&id=20&Itemid=25), а также ранжирования сайтов ОНЗ РАН, проведенного в Дальневосточном геологическом институте ДВО РАН (Табл. 1).

Пользователь сервиса может осуществить ранжирование сайтов самостоятельно, выбрав необходимое ему множество сайтов, признаки и формулу. Результаты многомерного анализа сайтов ОНЗ РАН можно также посмотреть на сервисе http://webometrics.fegi.ru/index.php?option=com_content&view=article&id=16&Itemid=15. Анализ признаков пространства научных сайтов ОНЗ РАН с использованием методов вебометрики и прикладной математики позволяет понять структуру и связность научных сайтов.

Рекомендации по позиционированию научных сайтов в Интернет

Кибернетическая лаборатория Центра научной информации и документации Национального исследовательского совета Испании на своем сайте опубликовала основные рекомендации по позиционированию научных сайтов в Интернет.

В работе [Ханчук и Наумова, 2009] также обсуждаются некоторые рекомендации по позиционированию научных сайтов. Использование этих рекомендаций позволит научному сайту увеличить доступность научной информации, опубликованной на данном сайте, для широкого круга потенциальных мировых пользователей.

Табл. 1. Рейтинг институтов Отделения наук о Земле Российской академии наук, февраль 2013 г.

| WR | Институт | S | V | R | Sc |
|----|--|----|----|----|----|
| 1 | Дальневосточный геологический институт ДВО РАН | 4 | 1 | 54 | 2 |
| 2 | Институт вулканологии и сейсмологии ДВО РАН | 2 | 3 | 1 | 3 |
| 3 | Институт географии РАН | 1 | 6 | 40 | 11 |
| 4 | Институт океанологии РАН им. П. П. Ширшова | 3 | 4 | 8 | 13 |
| 5 | Палеонтологический институт им. А. А. Борисяка РАН | 10 | 2 | 45 | 10 |
| 6 | Институт криосферы Земли СО РАН | 6 | 7 | 14 | 26 |
| 7 | Ордена Ленина и ордена Октябрьской революции Институт геохимии и аналитической химии им. В. И. Вернадского РАН | 7 | 11 | 17 | 7 |
| 8 | Минералогический музей им. А. Е. Ферсмана РАН | 19 | 5 | 44 | 20 |
| 9 | Институт нефтегазовой геологии и геофизики им. А. А. Трофимука СО РАН | 9 | 18 | 2 | 4 |
| 10 | Институт геологии и геохимии им. академика А. Н. Заварицкого УРО РАН | 25 | 30 | 6 | 1 |
| 11 | Институт экспериментальной минералогии РАН | 22 | 10 | 41 | 15 |
| 12 | Геофизическая служба РАН | 28 | 8 | 51 | 38 |
| 13 | Государственный геологический музей РАН им. В. И. Вернадского | 33 | 9 | 59 | 23 |
| 14 | Институт физики Земли им. О. Ю. Шмидта РАН | 50 | 12 | 9 | 9 |
| 15 | Лимнологический институт СО РАН | 37 | 17 | 3 | 12 |
| 16 | Институт геологии рудных месторождений, петрографии, минералогии и геохимии РАН | 16 | 13 | 48 | 17 |
| 17 | Институт геохимии им. А. П. Виноградова СО РАН | 20 | 23 | 7 | 14 |
| 18 | Институт солнечно-земной физики СО РАН | 26 | 27 | 5 | 18 |
| 19 | Институт горного дела им. Н. А. Чинакала СО РАН | 8 | 15 | 36 | 29 |
| 20 | Полярный геофизический институт Кольского НЦ РАН | 38 | 32 | 10 | 8 |
| 21 | Институт проблем нефти и газа РАН | 11 | 14 | 15 | 35 |
| 22 | Институт водных и экологических проблем СО РАН | 36 | 40 | 12 | 5 |
| 23 | Геологический институт РАН | 18 | 28 | 22 | 16 |
| 24 | Институт земной коры СО РАН | 14 | 26 | 11 | 30 |
| 25 | Институт оптики атмосферы им. В. Е. Зуева СО РАН | 54 | 16 | 58 | 6 |
| 26 | Институт озераведения РАН | 5 | 44 | 52 | 48 |
| 27 | Институт геоэкологии им. А. Е. Сергеева РАН | 17 | 21 | 29 | 28 |
| 28 | Институт геологии Уфимского НЦ РАН | 31 | 48 | 4 | 33 |
| 29 | Институт водных проблем РАН | 27 | 25 | 19 | 39 |
| 30 | Институт геологии Коми НЦ УрО РАН | 29 | 24 | 18 | 47 |
| 31 | Институт географии им. В. Б. Сочавы СО РАН | 32 | 36 | 28 | 24 |
| 32 | Международный институт теории прогноза землетрясений и математической геофизики РАН | 42 | 20 | 27 | 31 |
| 33 | Институт геологии и минералогии им. В. С. Соболева СО РАН | 30 | 35 | 16 | 36 |
| 34 | Тихоокеанский океанологический институт им. В. И. Ильичева ДВО РАН | 21 | 42 | 42 | 25 |
| 35 | Геологический институт КНЦ РАН | 39 | 19 | 26 | 40 |
| 36 | Институт природных ресурсов, экологии и криологии СО РАН | 23 | 50 | 13 | 32 |
| 37 | Северо-Восточный Комплексный научно-исследовательский институт ДВО РАН | 12 | 39 | 46 | 48 |
| 38 | Институт физики атмосферы им. А. М. Обухова РАН | 47 | 33 | 30 | 22 |
| 39 | Геологический институт СО РАН | 15 | 46 | 25 | 50 |
| 40 | Горный институт УрО РАН | 35 | 22 | 34 | 46 |
| 41 | Геофизический центр РАН | 44 | 31 | 20 | 41 |
| 42 | Тихоокеанский институт географии ДВО РАН | 51 | 45 | 47 | 21 |
| 43 | Институт тектоники и геофизики имени Ю. А. Косыгина ДВО РАН | 24 | 38 | 37 | 51 |
| 44 | Институт горного дела УРО РАН | 41 | 63 | 24 | 27 |
| 45 | Научная станция Российской академии наук в г. Бишкеке | 13 | 66 | 49 | 37 |
| 46 | Институт морской геологии и геофизики ДВО РАН | 43 | 29 | 50 | 42 |
| 47 | Институт геологии КарНЦ РАН | 34 | 41 | 31 | 44 |
| 48 | Институт экологических проблем Севера УРО РАН | 67 | 53 | 56 | 19 |
| 49 | Институт проблем комплексного освоения недр РАН | 49 | 43 | 21 | 47 |
| 50 | Институт мониторинга климатических и экологических систем СО РАН | 48 | 34 | 39 | 49 |
| 51 | Институт геофизики им. Ю. П. Булашевича УРО РАН | 45 | 47 | 23 | 45 |

Табл. 1. Рейтинг институтов Отделения наук о Земле Российской академии наук, февраль 2013 г. (продолжение)

| WR | Институт | S | V | R | Sc |
|----|---|----|----|----|----|
| 52 | Институт проблем промышленной экологии Севера КНЦ РАН | 58 | 37 | 38 | 34 |
| 53 | Горный институт КНЦ РАН | 40 | 51 | 55 | 58 |
| 54 | Байкальский институт природопользования СО РАН | 53 | 49 | 32 | 53 |
| 55 | Институт геологии алмаза и благородных металлов СО РАН | 63 | 57 | 33 | 43 |
| 56 | Институт геологии и геохронологии докембрия РАН | 61 | 54 | 35 | 55 |
| 57 | Институт угля СО РАН | 56 | 52 | 60 | 57 |
| 58 | Институт минералогии УрО РАН | 52 | 61 | 65 | 54 |
| 59 | Институт динамики геосфер РАН | 57 | 60 | 61 | 49 |
| 60 | Институт водных проблем Севера КарНЦ РАН | 46 | 64 | 65 | 58 |
| 61 | Институт водных и экологических проблем ДВО РАН | 62 | 55 | 43 | 58 |
| 62 | Институт горного дела ДВО РАН | 64 | 56 | 53 | 56 |
| 63 | Институт горного дела Севера им. Н. В. Черского СО РАН | 55 | 65 | 57 | 57 |
| 64 | Институт мерзлотоведения им. П. И. Мельникова СО РАН | 65 | 58 | 62 | 52 |
| 65 | Ильменский государственный заповедник им. В. И. Ленина УрО РАН | 60 | 62 | 63 | 60 |
| 66 | Институт степи УРО РАН | 59 | 67 | 64 | 58 |
| 67 | Специальное конструкторское бюро средств автоматизации морских исследований ДВО РАН | 66 | 59 | 65 | 59 |

Литература

- Мазалов, В. В., и А. А. Печников (2009), О рейтинге официальных сайтов научных учреждений северо-запада России, *Управление большими системами, Выпуск 24*, ИПУ РАН, Москва, 130–146.
- Наумова, В. В. (2012), Мониторинг Информационного пространства Дальневосточного отделения РАН, *Вестник ДВО РАН*, 2, 117–122.
- Печников, А. А. (2008), Вебометрические исследования Web-сайтов университетов России, *Информационные технологии*, 11.
- Печников, А. А. (2010), Модель университетского Веба, *Вестник Нижегородского университета им. Н. И. Лобачевского*, 6, 208–214.
- Ханчук, А. И., и В. В. Наумова (2009), Информационное пространство Дальневосточного отделения РАН, *Вестник ДВО РАН*, 4, 122–130.
- Шокин, Ю. И., О. А. Клименко, Е. В. Рычкова, И. В. Шабальников (2008), Рейтинг сайтов научных организаций СО РАН, *Вычислительные технологии*, 13, 3, 128–135.
- Aguillo, I. F., B. Granadino, J. L. Ortega, J. A. Prieto (2006), Scientific research activity and communication measure with cybermetrics indicators, *Journal of the American Society for the Information Science and Technology*, 57, 10, 1296–1302, doi:10.1002/asi.20433.
- Almind, Tomas C., and Peter Ingwersen (1997), Informetric analyses on the World Wide Web: Methodological approaches to “webometrics”, *Journal of Documentation*, 53, 4, 404–426, doi:10.1108/EUM0000000007205.

В. В. Наумова и К. А. Платонов, Дальневосточный геологический институт ДВО РАН, проспект 100-летия Владивостока 159, 690022, г. Владивосток. (naumova@fegi.ru)