



Тезисы доклада

1. НАЗВАНИЕ ДОКЛАДА:

Построение рейтинга вузов по данным Википедии

Building a world ranking of universities according to the data from Wikipedia

2. АВТОРЫ:

Ярышкина Е. А.

Yaryshkina E. A.

3. ОРГАНИЗАЦИЯ (полное наименование, без аббревиатур):

Петрозаводский государственный университет

Petrozavodsk State University

4. АННОТАЦИЯ:

Существует большое разнообразие национальных и международных университетских рейтингов. Составители рейтингов ориентируются на различные параметры, оценки качества образования, статьи и т.д. Для построения нового рейтинга вузов может выступать Википедия, свободная энциклопедия, где каждое значительное учебное заведение имеет свою страницу. Основная задача при построении рейтинга — выбрать надежные свойства у страниц Википедии, механизм построения рейтинга и проверить насколько результаты нового рейтинга согласуются с глобальными университетскими рейтингами.

There are a great variety of national and international university rates. Ratings are oriented to different characteristics, education quality control, number of written articles etc. Every significant institution has its own web page in Wikipedia. Information extracted from this Wikipedia page can help to form a new rating. To choose reliable qualities on Wikipedia pages, mechanism of rate building and to check how much new rate's results conform with global university rates are the main goals of this research.

5. КЛЮЧЕВЫЕ СЛОВА:

Википедия, рейтинг, вебометрика

Wikipedia, rating, webometrics

6. ТЕКСТ ТЕЗИСОВ ДОКЛАДА:

Для современного развивающегося государства высокий уровень качества образования является важным фактором, который определяется лидирующими позициями ведущих национальных университетов на международном уровне. Сейчас ФАНО проводит конкурс по научно-технологическому прогнозированию с целью анализа глобальных трендов развития науки и технологий [1].

Высокие позиции в наиболее популярных международных университетских рейтингах можно считать визитной карточкой для любого университета, поскольку все больше абитуриентов внутри страны и за рубежом ориентируются на рейтинги.

Традиционные рейтинги опираются на различные показатели учебной и научной деятельности преподавателей, выпускников и сотрудников вузов. При этом на рейтинги влияют не только объективные количественные показатели (количество публикаций, ссылок и престижных международных наград, например Нобелевской премии), но и различные субъективные мнения – отзывы выпускников, работодателей и т.п.

История традиционного академического ранжирования университетов начинается с 1870 года, когда Комиссия Бюро по образованию США опубликовала первый ежегодный отчет со статистической информацией и классификацией учебных заведений. В 1983 году начался второй этап истории университетских рейтингов — журнал «US News&World Report» опубликовал список 50-ти лучших вузов США, созданный в помощь абитуриентам и их родителям и инициировавший в дальнейшем систематическую работу в этом направлении. С 2003 года начался третий этап в развитии практики ранжирования вузов и связан он с первым шагом в построении системы глобальных университетских рейтингов [2]. Опубликованный рейтинг 500 ведущих вузов мира (ARWU) Института высшего образования Шанхайского университета можно считать началом эпохи мировых рейтингов высшего образования.

В 2004 году был запущен испанский проект Webometrics [3] — глобальный рейтинг, формируется лабораторией «Cybermetrics», входящей в исследовательскую группу Центра информации и документации Национального Исследовательского Совета Испании. Cybermetrics Lab оценивает присутствие вузов в виртуальном информационном пространстве и опирается в основном на Интернет-показатели университетских сайтов. Если принять гипотезу вебометрики, что рейтинг сайтов

соответствует значимости действительных объектов, можно построить рейтинг вузов. Лаборатория Cybermetrics Lab разработала методологию и алгоритм ранжирования веб-сайтов на основе четырех показателей: видимость — число внешних ссылок на страницы сайта вуза; размер — число страниц сайта; специальные файлы — количество PDF-, DOC- и PPT-файлов; научные публикации — число публикаций сайта и ссылок на них других авторов по данным Google Scholar.

У каждого значительного вуза есть статья в Википедии. Цель исследования заключается в том, чтобы определить: какие параметры Википедии являются надежными (стабильными, то есть медленно изменяются во времени), и соответственно, пригодными для построения рейтинга статей Википедии. Надёжность параметра заключается в отсутствии непредвиденных недопустимых изменений в процессе изменения статьи Википедии. От рейтинга статей можно перейти к рейтингу вузов, если принять гипотезу викиметрики (wikiometrics) — степень популярности статей в Википедии соответствует позиции объекта (здесь — вуза) в рейтинге. Викиметрика — новая концепция [4]. Для первого этапа нашего исследования было выбрано 65 статей из Английской Википедии, которые были разделены на две группы. Статья может иметь статус «избранной», «хорошей» либо «добротной» статьи или не иметь статуса вовсе. Первая группа — избранные и хорошие статьи, вторая группа — статьи без статуса.

В ходе исследования рассматриваются следующие параметры статей Википедии:

1. Статус статьи;
2. Объем статьи (в килобайтах);
3. Число редакторов;
4. Число правок;
5. Число интервики — это число гиперссылок, ведущих на статьи Википедии на других языках о данном объекте;
6. Число просмотров;
7. Количество наблюдателей;
8. Количество перенаправлений;

Перенаправление — это обычная страница Википедии, текст которой содержит команду перенаправления на статью с другим названием. Пример: чтобы перенаправить пользователя, открывающего страницу “МГУ”, на страницу “Московский государственный университет”, необходимо на странице “МГУ” указать: `#REDIRECT [[Московский государственный университет]]`. Теперь, если пользователь

напишет в строке поиска “МГУ”, то будет перенаправлен поисковой системой сайта Википедии на страницу “Московский государственный университет”.

9. Число входящих ссылок;
10. Число исходящих ссылок;
11. Число внешних ссылок.

Соотношения между этими параметрами и рейтингом университетов (по данным Cybermetrics Lab) анализируются для того, чтобы предсказать рейтинг университета, используя только данные из Википедии.

Статьи разделены на группы для того чтобы проанализировать параметры статей и определить:

1. «Плохие» или «нестабильные» параметры, которые могут быть изменены одним человеком (или небольшой группой редакторов);
2. «Хорошие» или «стабильные» параметры, которые являются результатом действий большого количества людей (редакторов и читателей Википедии). Именно эти «стабильные» параметры Википедии являются надежными, медленнее изменяются во времени и могут быть использованы для построения рейтинга вузов на основе данных Википедии.

Для выявления связей между параметрами был рассчитан линейный коэффициент корреляции Пирсона, таким образом, были выявлены пары взаимосвязанных параметров:

- число просмотров и число входящих ссылок;
- число входящих ссылок и количество редакторов;
- число редакторов и количество правок;
- число внешних ссылок и объем страницы.

Однако чтобы использовать эти взаимосвязи в дальнейшем, необходимо доказать что распределение параметров является нормальной величиной. Эта задача осложняется тем, что в нашем исследовании каждая статья Википедии имеет 11 параметров.

В дальнейшем для ранжирования университетов можно применить такой метод как система голосования, где каждый из параметров получит свой вес. При построении итогового рейтинга следует использовать только значимые параметры, которые влияют на результат и место каждого вуза в рейтинге.

В данном исследовании мы планируем выстроить страницы вузов в Википедии по рангу. Существует метод, который может быть адаптирован и применен для построения рейтинга вузов — алгоритм ссылочного ранжирования PageRank. PageRank — это метод Google для измерения «важности» страницы. Так же как Google

использует PageRank, чтобы откорректировать результаты так, что более «важные» сайты поднимутся вверх на странице результатов поиска пользователя, мы можем применить алгоритм для поиска «важных» вузов. Для применения PageRank к вузам, построим ссылки между вузами (точнее их страницами в Википедии) следующим образом. Для каждого из 11 параметров выстраиваем значения в порядке убывания — самое большое число занимает первое место, следующее за ним занимает второе место и т.д. Страница, параметр которой занял первое место, получает «ссылки» от всех других страниц. Страница, параметр которой занял второе место, получает «ссылки» от всех других страниц кроме той, на которую ссылается. Последняя по рангу страница не получит ни одной «ссылки». К построенному графу можно применить классический алгоритм PageRank.

После построения рейтинга вузов по данным из Википедии его следует сравнить уже с существующими глобальными рейтингами, такими как ARWU, THE и Webometrics. Если не будет значительных расхождений между новым рейтингом и мировым рейтингом, то можно сделать вывод о том что рейтинг построенный по данным из Википедии имеет место для существования, степень популярности статьи соответствует позиции вуза в рейтинге и можно прогнозировать положение других вузов. Это важно для оценки многих российских вузов, не входящих в мировые рейтинги. В Википедии на 2015 год существуют страницы 780 российских вузов.

Список литературы:

1. Конкурс центров научно-технического прогнозирования.— Электронный ресурс. URL: http://fano.gov.ru/ru/activity/Competitions/Comp_tsntp/index.php
2. Аноприенко А.Я. Университет в современном информационном пространстве: тенденции, рейтинги и опыт развития портала магистров ДонНТУ // Научные труды ДонНТУ: журнал.—2011.—№ 13(185).—С.224-235.—ISSN 1996-1588.
3. Клименко О.А., Петров И.С. Исследование строения и динамики развития научного веб-пространства на примере СО РАН // Тезисы докл. XIII Росс. конф. с участием иностранных ученых «Распределенные информационные и вычислительные ресурсы» (DICR'2010) (Новосибирск, Россия, 30 ноября-3 декабря 2010). – № гос. регистр. 0321100051, ФГУП НТЦ «Информрегистр». – Новосибирск. – 2010. – С. 35.
4. Rokach L., Katz G. Wikiometrics: A Wikipedia Based Ranking System // ResearchGate, January. – 2014.