

# Рунет перевалил за миллиард!

В ответ на вопросы специалистов об объемах "базы" интеллектуальной поисковой системы Nigma.ru, участники научного проекта Nigma.ru провели оценку суммарного объема русскоязычных документов в поисковых системах, с помощью которых в данный момент осуществляет поиск проект Nigma.ru, и получили следующий результат: по состоянию на 16-е мая 2005 года в Интернете имеется примерно 1'052'227'229 русскоязычных документов.

Это вторая попытка группы посчитать размер Рунета — первая окончилась неудачно, т.к. предложенный тогда алгоритм не обладал т.н. "устойчивостью" — при изменении некоторых параметров, оценка количества документов не колебалась около т.н. "равновесного решения", а резко изменялась. Новый алгоритм лишен этого недостатка — он основывается на "закачивании" случайных запросов, созданных на базе частотного словаря, и сравнении общего объема выданных результатов со всех поисковых машин, с которыми работает Nigma.ru, и эталонной поисковой машины, о которой известен размер ее базы.

Первоначально, устранение дублей происходило по URL и заголовкам документов, однако во второй версии, с использованием частотного словаря, количество дублей резко сократилось, и сейчас используется только URL. Если исключить из алгоритма одну из поисковых машин, оценка в среднем падает всего на 12%. Соответственно, возможно, что не более 20-30% документов Рунета осталось непроиндексированными ни одной из поисковой машин, использующихся в проекте Nigma.ru.

Не совсем корректно называть полученное число как объем "базы" поисковой системы Nigma.ru, т.к. собственной базы у проекта нет. Вместо этого предлагается воспринимать это число как оценку общего количества русскоязычных документов в Интернете.

Участники проекта надеются, что это исследование даст толчок многим поисковым системам и, в результате, российский Интернет будет проиндексирован еще лучше и глубже, чем он проиндексирован сейчас, что даст возможность пользователям Рунета искать и находить важную для них информацию гораздо быстрее и проще.

Автор: Артур Скальский © Babr24.com ИНТЕРНЕТ , МИР 👁 2161 17.05.2005, 12:57 📌 226

URL: <https://babr24.com/?ADE=21732> Bytes: 2019 / 2019 Версия для печати

 [Порекомендовать текст](#)

Поделиться в соцсетях:

*Также читайте эксклюзивную информацию в соцсетях:*

- [Телеграм](#)

- [ВКонтакте](#)

*Связаться с редакцией Бабра:*

[newsbabr@gmail.com](mailto:newsbabr@gmail.com)

## НАПИСАТЬ ГЛАВРЕДУ:

Телеграм: [@babr24\\_link\\_bot](#)

Эл.почта: [newsbabr@gmail.com](mailto:newsbabr@gmail.com)

## ЗАКАЗАТЬ РАССЛЕДОВАНИЕ:

эл.почта: [bratska.net.net@gmail.com](mailto:bratska.net.net@gmail.com)

## КОНТАКТЫ

Бурятия и Монголия: Станислав Цырь  
Телеграм: @bur24\_link\_bot  
эл.почта: bur.babr@gmail.com

Иркутск: Анастасия Суворова  
Телеграм: @irk24\_link\_bot  
эл.почта: irkbabr24@gmail.com

Красноярск: Ирина Манская  
Телеграм: @kras24\_link\_bot  
эл.почта: krasyar.babr@gmail.com

Новосибирск: Алина Обская  
Телеграм: @nsk24\_link\_bot  
эл.почта: nsk.babr@gmail.com

Томск: Николай Ушайкин  
Телеграм: @tomsk24\_link\_bot  
эл.почта: tomsk.babr@gmail.com

[Прислать свою новость](#)

#### **ЗАКАЗАТЬ РАЗМЕЩЕНИЕ:**

---

Рекламная группа "Экватор"  
Телеграм: @babrobot\_bot  
эл.почта: equatoria@gmail.com

#### **СТРАТЕГИЧЕСКОЕ СОТРУДНИЧЕСТВО:**

---

эл.почта: babrmarket@gmail.com

[Подробнее о размещении](#)

[Отказ от ответственности](#)

[Правила перепечаток](#)

[Соглашение о франчайзинге](#)

[Что такое Бабр24](#)

[Вакансии](#)

[Статистика сайта](#)

[Архив](#)

[Календарь](#)

[Зеркала сайта](#)