

УДК 378.147

## ВЕРИФИКАЦИЯ ИНТЕРНЕТ - ИНФОРМАЦИИ

И.Ю. Анохина,

Донецкий национальный технический университет  
IngaTula@mail.ru

*Анохина И.Ю. Верификация Интернет – информации. Рассмотрены вопросы верификации информации, получаемой из Интернета. Показано, какими способами и средствами можно проверить сайты, определить их направленность, получить информацию об авторе размещенного материала, оценить степень плагиата в статье, книге.*

### Введение

В наше время поиск информации в Интернете стал обыденностью. К нему прибегают, если нужно узнать сводку погоды, получить информацию по бытовым вопросам, прочитать новости, научные статьи, ознакомиться с лекциями коллег, найти художественную и научную литературу.

В ряде случаев источник информации особой роли не играет. Узнавая новые рецепты или отыскивая художественную литературу, мы можем не беспокоиться о том, с какого сайта была взята информация. Но есть, по крайней мере, два направления, которые требуют тщательной подборки сайтов, откуда будет считываться информация. Это политические новости и научные исследования.

Если в первом случае можно в какой-то мере определиться с позициями участников и читать новостные порталы тех, с кем эти позиции совпадают, то в случае получения информации по профессиональной сфере деятельности, точного ответа, какому сайту можно доверять, а какому нет, не существует. В качестве совета обычно используется фраза «нужно брать информацию только с проверенных сайтов». Вопрос в том, где найти эти сайты, кто их проверял и действительно ли они являются источниками точной информации, а не преобразованной в соответствии с желаниями авторов.

С точки зрения использования Интернета в профессиональной сфере, следует учитывать, что Интернет в научных исследованиях используется не только, как источник информации, но и как средство ее обработки (on-line ресурсы), а также как средство коммуникации.

Говорят, что в Интернете можно найти ответ на любой вопрос. Это правильно. Но будет ли ответ верным. Являясь огромным, постоянно наполняющимся хранилищем информации, Интернет имеет существенный недостаток. Автором может стать каждый, что с высокой степенью вероятности приводит к появлению

недостовверной, непроверенной информации или плагиата. Любой пользователь Интернета может разместить свою статью, например, в Википедии [1], в любом журнале, в блоге, живом дневнике. Качество информации, как правило, не проверяется, за исключением сайтов государственных организаций, правительственных и пр.

Компания Netcraft, проводящая исследования в области Интернета, в сентябре 2014г. сообщила о том, что количество сайтов в Интернете превысило миллиард (на май 2015г. их количество сократилось до 857 927 160) [2].

С одной стороны такое огромное количество сайтов, безусловно, является достоинством, так как чем больше информации, тем легче ее проверить, сравнить, проанализировать, но с другой стороны обилие сайтов приводит к затрудненному поиску информации, когда просто более сильно раскрученные сайты не претендующие на точность информации, а берущие лишь ее количеством, оказываются на первых страницах поисковых систем, обгоняя академические сайты.

Качество информации трудно верифицировать, нет гарантии, что скачанная статья, подписанная известным специалистом, на самом деле принадлежит ему или излагаемый в ней материал не претерпел изменений при переводе или не интерпретирован отлично от авторского.

Классифицируя источники информации Интернета, можно выделить несколько категорий:

#### **1. Категории сайтов с высоким уровнем достоверности информации:**

- сборники, монографии, изданные по рекомендациям ученых советов;
- сайты известных ученых, само имя которых является гарантией качества информации;
- сайты кафедр вузов;
- изданные академические словари, энциклопедии.

#### **2. Категория сайтов с непроверенной**

**информацией:**

–сайты электронных энциклопедий (Википедия и аналогичные);

–реферативные сайты [3,4].

Наполняемость последних двух видов сайтов осуществляется за счет широкого участия всех желающих, причем далеко не у всех из них наличествует высшее образование.

Если требуемая информация не была найдена на сайтах первой категории, желательна проверка сайтов, откуда будут считываться материалы.

Ярким примером такого вида сайтов является сайт **Новости технологий** (рис.1, цифра 1), публикующий подборки статей о новейших достижениях науки и техники [5].

The image shows a composite of three parts related to the website 'Новости технологий' (Tech News):

- Section 1:** The top part of the website, including the title 'НОВОСТИ ТЕХНОЛОГИЙ', a quote about the invention of the ballpoint pen, and navigation links like 'Главная', 'Видеоролики', 'Фотогалерея', and 'Обзоры'. It also lists the editor 'Tech' and authors 'Технолог' and 'Акеда'.
- Section 2:** A 'Контакты' (Contacts) section with a form to check the website address. Below it is a WHOIS record for 'techvesti.ru', showing it is a private person registered in RU-CENTER-RU on 2008-11-02.
- Section 3:** A search interface on Yandex.ru. A search for an image from the website results in a list of sites where the image is found, including 'Новый энергетический проект Lockheed Martin' and 'Top-25 изобретений уходящего года'.

Рисунок 1 – Анализ сайта и поиск исходной статьи

Несмотря на широкую рекламу, достаточно высокую посещаемость, участие в различных видах рейтингов, вряд ли можно считать публикуемую на нем информацию стопроцентно достоверной. На сайте размещаются в основном переводные статьи о новостях науки и техники.

В качестве контактов редакции сайта указаны авторы Технолог и Акеда (рис.1, цифра 1). Вряд ли использование вымышленных имен способствует возрастанию доверия к сайту. В редакции не указаны имена ученых, таким образом, подборку, перевод статей осуществляют неизвестные люди, об уровне компетентности которых ничего неизвестно. Если учесть, что перевод технического текста требует знания профессионального словаря и сленга, то нет гарантии, что перевод не был выполнен без ошибок.

Сайт можно проверить для определения на чьи имя и где он зарегистрирован.

Таких систем проверки существует несколько, одним из них является сервис **СУ-**

**PR.com**, осуществляющий анализ сайта по его адресу (рис.1, цифра 2) [6].

После ввода адреса анализируемого сайта выдается информация о владельце, дате и месте регистрации, списке ключевых слов, используемых при поиске. Как показала проверка, сайт зарегистрирован на анонимное лицо в Петербурге. Какой вывод следует сделать? Сайт можно использовать для чтения новостей, однако, гарантии точности информации нет.

Может получиться так, что статья, размещенная на сайте, нужна для работы. В таком случае желательно найти оригинал.

Способов поиска оригиналов статьей много, самым простым является поиск по рисункам, если они есть.

Для проверки открывают статью, находят любой используемый в ней рисунок, сохраняют его на компьютере и обращаются к поиску в Яндексе. Эта поисковая система позволяет осуществлять поиск по картинкам (терминология Яндекса). На Яндекс загружают картинку из статьи и поисковик ищет аналоги.

На рис.1 цифрой 3 показан описанный

поиск по картинке. Как видно, найдено несколько сайтов, содержащих используемую картинку, в том числе и сайт фирмы-производителя термоядерного реактора, о котором велась речь в статье. Это оригинальная статья, именно ее и следует использовать как источник информации.

Аналогичный поиск по изображению реализован в Google, результаты обоих поисковиков не всегда совпадают, поэтому желательно вести поиск параллельно.

Способ поиска по картинке можно использовать для проверки на авторство статьи, работ студентов, докладов на конференциях...

При анализе сайта сервисом **CY-PR.com**, как впрочем, и любым другим, как показали

многочисленные проверки, возможны ошибки.

Например, специализированный каталог научных конференций «Конференции.ру» (<http://www.konferencii.ru>), публикует данные о прошедших и предстоящих конференциях в России, Беларуси и других странах СНГ (цифра 1, рис.2).

При проверке с помощью сервисов **CY-PR.com** [6] и <http://www.seogadget.ru/> [7] были выявлены несоответствия в результатах. Так, по данным **CY-PR**, сервер, на котором размещен сайт, находится в Беларуси (цифра 2, рис.2, подчеркнуто), а по данным сервиса **Seogadget** – в Москве (цифра 3, рис.2, подчеркнуто).

1. **Конференции.ру** <http://www.konferencii.ru>  
 8 июня 2015 г. — 11 июня 2015 г., срок заявок: 20 апреля 2015 г. (приём заявок закончен)  
**Международная Российско-Казахстанская школа-конференция «Химически**  
 Конференция | Естественные науки, Материаловедение, Молодые ученые, Нанотехнологии, Углерод  
 Россия, Новосибирск  
 Организаторы: Новосибирский государственный технический университет

2. <http://www.cy-pr.com>  
 9 июня 2015 г. — 11 июня 2015 г., срок заявок: 15 мая 2015 г. (приём заявок закончен)  
**Международный форум «Ветеринарная медицина и продовольственная безопасность**  
 Форум | Ветеринария  
 Россия, Ульяновск (ИИЧ)  
**[www.konferencii.ru](http://www.konferencii.ru) | WHOIS**  
 Владелец: Private Person  
 Регистратор: REGTIME-RU  
 Статус: VERIFIED  
 Дата регистрации: 2007-06-21 [2913 дня назад]  
 Дата окончания: 2015-06-21 (через 9 дней)  
 Заголовок: Конференции.RU | Каталог научных мероприятий  
 Описание: Открытый каталог научных конференций, выставок и семинаров выставки заочные конференции международные конференции  
 Ключевые слова: конференции научные конференции конференции семинары выставки заочные конференции международные конференции  
 Статус о сайте: НЕТ Заказать  
 IP адрес сервера: 178.159.253.111  
 Сервер расположен в: Беларусь

3. <http://www.seogadget.ru/>  
 URL: [www.konferencii.ru](http://www.konferencii.ru)  
 IP адрес: 178.159.253.111 (Россия / Москва)  
 Регистрация домена: 21.06.2007  
 Домен оплачен до: 21.06.2016  
 Открывается ли сайт: ДА

4. **you get signal**  
**Visual Trace Route Tool**  
 approximate geophysical trace  
 Host trace to: 178.159.253.111  
 14 hops / 6.3 seconds  
 1. 107.170.234.254  
 2. 198.199.99.237  
 3. sfmxx.org  
 4. he.net  
 5. he.net  
 6. he.net  
 7. he.net  
 8. 216.66.87.134  
 9. 95.167.93.162  
 10. dtn.ru  
 11. Unknown  
 12. 95.131.30.186  
 13. activeby.net  
 14. activeby.net  
 ~5,920 miles traveled  
 Remote Address: 178.159.253.111 Host Trace Proxy Trace

Рисунок 2 – Сравнение результатов анализа сайтов на разных сервисах

В таких случаях желательна дополнительная проверка.

Одним из лучших, по мнению автора, сервисов является сервис **You get signal - Get Trace Route Tool** [8], позволяющий не только анализировать, но и визуализировать местоположение сервера.

На рис. 2 цифрой 4 отмечен скрин экрана сервиса с визуализацией местоположения сервера. Эта часть функций сервиса является бесплатной. Кроме этого есть функции платные: определение данных (фамилии авторов, местоположения, аккаунтов в социальных сетях, выложенных фотографий и видео) по адресу электронной почты. Стоимость варьируется от 1\$ до 40\$ в зависимости от количества информации.

С помощью описанных сервисов можно получить информацию о сайте, но еще одним важным аспектом проверки информации, является проверка персоны, разместившей эту информацию в сети.

Для примера выбран автор немецкоязычного сервера Исследовательского центра в Карлсруэ (цифра 1, рис.3), специализирующегося на разработках в области естественных и инженерных наук (<http://www.kit.edu/index>).

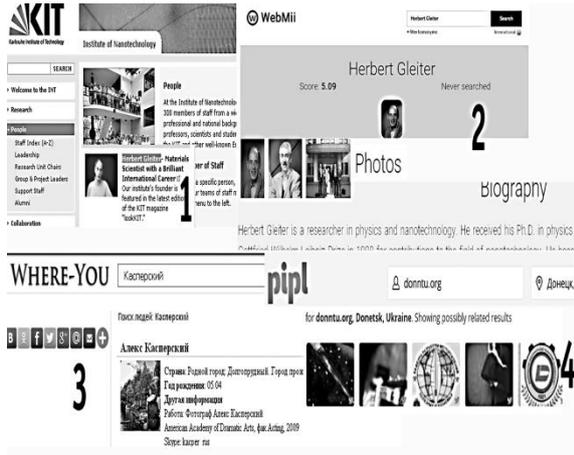


Рис 3 – Поиск человека по аккаунтам в сетях, адресу электронной почты, фамилии

Одним из популярных сервисов является сервис **WebMii** [9], осуществляющий поиск или по имени, или по набору ключевых слов, описывающих личность. Как видно из рис.3 (цифра 2) в качестве результата получены фотографии, биография и профиль в социальных сетях, через которые можно связаться с ученым. Отметим, что поиск позволяет искать и в русскоязычной аудитории, однако результатов значительно меньше.

Отличным сервисом поиска по социальным сетям среди русскоязычного населения является сервис **Where-you** (<http://where-you.com>). Он позволяет найти профили в Контакте, Одноклассниках, Фейсбуке, Google, Твиттере, Моем Мире. К недостаткам следует отнести неоптимальный алгоритм поиска. Так, по запросу «Александр Касперский» было найдено 2700

аккаунтов, причем в ряде из них совпадали или имя, или фамилия. А при попытке найти человека по адресу электронной почты было найдено более четырех миллионов человек, имеющих почту mail.ru (в условиях поиска была задана именно она).

Сравнивая оба сервиса, можно рекомендовать сервис **WebMii** для поиска англоязычных авторов как более информативный и резко ограничивающий количество результатов, для русскоязычной аудитории - сервис **Where-you** несмотря на большое количество результатов, несоответствующих условиям поиска.

Сервис **Pipl** [10] осуществляет поиск по имени, фамилии, адресу электронной почты и адресу сайта. На рис. 3. цифрой 4 отмечен поиск, выполненный по запросу «donntu.org» - адрес официального сайта ДонНТУ. На основании поиска открываются аккаунты вуза в социальных сетях, что доказывает их необходимость для распространения информации об учебном заведении.

Достоинством сервиса **Pipl** является то, что в отличие от ранее описанного сервиса **Where-you**, он позволяет осуществлять поиск по имени, фамилии, адресу электронной почты, городу и стране как по совокупности условий, что отсекает лишние результаты.

Еще один популярный сервис <http://socsearcher.net> (цифра 1, рис.4), который, во-первых, имеет хорошо организованный поиск по русскоязычным сайтам, а во-вторых, отсекает ненужную, несоответствующую поставленному запросу, информацию.

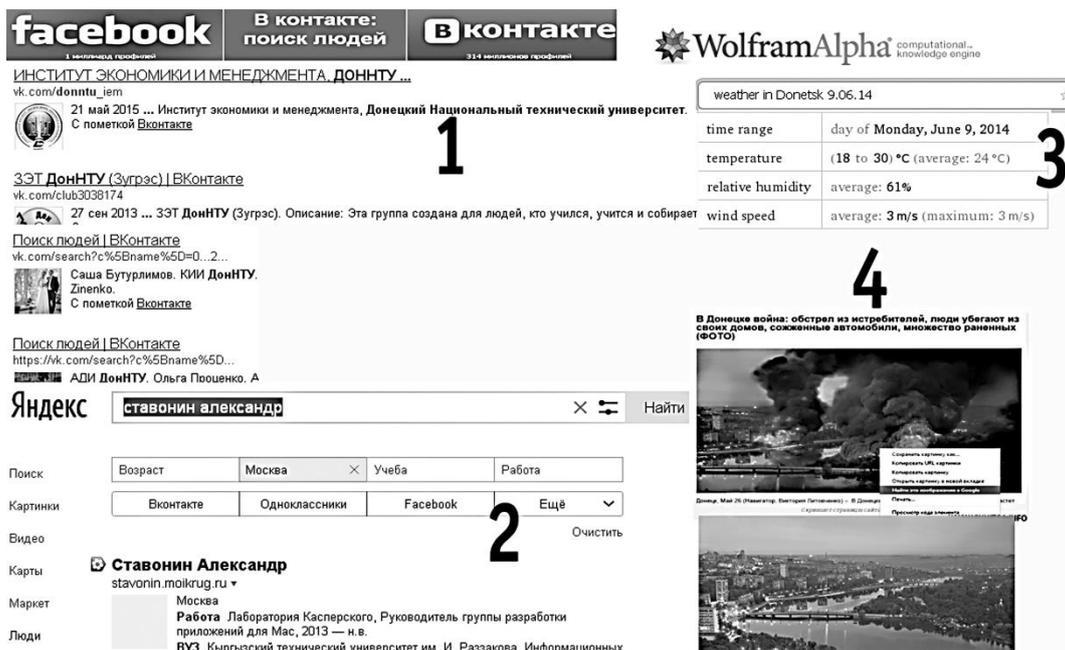


Рис 4 – Поиск аккаунтов и проверка фотографий на подлинность

Как показали исследования, ДонНТУ отлично представлен на этом сервере (148000 результатов). Это связано в частности с большим количеством аккаунтов, созданных студентами и выпускниками вуза.

Сервис, реализующий дополнительные возможности по отношению к ранее рассмотренным - **Wolfram Alpha**, определяемый как интеллектуальный робот, выдающий любую информацию, ориентируясь на энциклопедии, официальные сайты и пр. Недостатком системы является невозможность распознавания кириллицы, запросы вводятся на английском языке.

И **Google** и **Yandex** позволяют не только искать людей в социальных сетях (цифра 2, рис.4), но часто используются для определения фейков, что особенно актуально для новостных порталов. Цифрой 3 отмечена фотография Донецка с изображением большого пожара. С помощью «поиска по картинке» **Google** (сервис аналогичный рассмотренному ранее на Яндекске)

удалось найти оригинальное изображение, на которое впоследствии было наложено изображение пожара.

Этот способ позволяет наравне с проверкой изображений в научных статьях, осуществлять проверку информации, выкладываемую на новостных порталах.

Еще одной проблемой, которую принес Интернет, является плагиат. Это может быть и использование чужих материалов при написании курсовых работ студентами, и вкрапления чужих текстов в научных работах.

Как и в случае поиска информации, вариантов много. Можно рекомендовать популярный сервис <http://text.ru/antiplagiat/>, осуществляющий проверку текста на уникальность [11].

В качестве материала для проверки взяты новости на дату написания статьи (27.06.15г.) Предполагалось, что, так как это свежие новости, текст должен быть уникален.

Результаты проверки показаны на рис.5.

The screenshot shows the TEXT.RU interface. At the top, it says 'Проверка текста на уникальность' and 'Проверка орфографии'. The main results show a uniqueness score of 1.40% and 2 orthography errors. Below this, there is a list of domains with their uniqueness scores, all at 98%. A detailed view of a text snippet is shown, with a uniqueness score of 1% for a specific fragment. The interface also includes a 'Сложность текстов' section and a 'Версии текста' section.

Рисунок 5 – Проверка текста на подлинность

Цифрой 1 отмечен исходный текст. Как видно из результатов проверки, уникальность текста в первый раз как исходного материала оценивалась в 1%( цифра 2 рисунка). Цифрой 3

на рисунке отмечены портала, имеющие совпадение на уровне 98%. Оценка времени появления информации позволила сделать вывод, что новостной портал РИАновости выдал информацию первым, остальные только,

пользуясь терминологией социальных сетей, перепостили материал.

Кроме этого при анализе текста приводятся такие характеристики, как «заспамленность» текста и «вода». «Заспамленность» предполагает наличие скрытой рекламы и количества поисковых ключевых слов в тексте. Под «водой» понимают фрагменты текста, не несущие никакой смысловой нагрузки, использование стоп-слов, устоявшихся фразеологизмов. Если результат оценки уровня «воды» менее 15%, то текст считается корректным. Оценка воды на уровне 10% (цифра 2 рисунка) позволяет утверждать, что именно на столько процентов текст можно сократить, при этом смысл не потеряется.

Допустимым уровнем «заспамленности» считается 30% [12].

При дополнительной проверке, проведенной на этом же сервисе, было выявлено два источника, которые имеют одинаковый текст. Оценить, какой из источников является оригиналом, а какой – плагиатом или, как в данном случае, цитированием, можно по дате и времени выдачи информации.

Цифрой 4 на рис.5 отмечены домены, имеющие совпадение текста. Сервис позволяет получить не только количественную характеристику совпадений, но и увидеть, какие именно фрагменты совпадают (цифра 5 рис.5).

Аналогичным образом был проверен фрагмент из автореферата диссертации (цифра 6 рис.5). Было выявлено еще два портала, на которых эта же диссертация была выложена, но т.к. автор был один и тот же, текст следовало считать уникальным.

Автор проверил на уникальность аннотацию этой статьи. Как показала проверка, уникальность составляла 100%.

Следовательно, сервис можно использовать и для проверки собственных документов с целью выявления процента неинформативности.

## Литература

1. Википедия. Свободная энциклопедия. Режим доступа: <https://ru.wikipedia.org>
2. Сайт компании Netcraft. Режим доступа: <http://news.netcraft.com/archives/category/web-server-survey/>
3. Ильин К.Ценность источников информации. Information Security/ Информационная безопасность № 1-2, 2006. – с.44-45.
4. Омелин М. Матрица достоверности информации для общедоступных on-line текстов, используемых в ходе научных исследований. Режим доступа:

## Выводы

В статье рассмотрены вопросы верификации информации, получаемой из Интернета. Проверка на подлинность, уникальность, правдивость информации необходима, как при поиске материалов для научных исследований, так и при чтении новостных лент, документальной литературы и пр.

Для установления уровня достоверности информации рекомендуется проверять сайты, на которых информация была размещена. Рассмотрены возможные варианты такой проверки. С учетом того, что при проверке возможно получение ошибочных результатов, рекомендовано проверять сайты на разных серверах и сравнивать результаты.

Кроме этого, для того, чтобы убедиться в достоверности информации, желательно проверить и личность человека, выложившего ее в сеть. Способы проверки описаны и проанализированы.

При работе желательно иметь оригинальное произведение, а не плагиат. Показана работа сервиса, осуществляющего проверку на уровень уникальности текста. Кроме уровня уникальности выдаются еще две характеристики :оценка уровня наличия спама и «воды».

Сервис может быть полезен и авторам, так как такая характеристика, как «вода» заставляет задуматься о наличии лишних, не несущих информацию слов в тексте.

В целом следует отметить, что наличие в Интернете огромного количества непроверенной информации заставляет выработать стратегию ее проверки.

<http://omelin.ru/index.php/publikatsii/25-matrix-of-reliability> - Авторский портал Михаила Омелина.

5. Новости технологий. Режим доступа: <http://techvesti.ru>.

6. Сервис CY-PR.com. Режим доступа CY-PR.com.

7. Seogadget. Инструменты веб-мастера. Режим доступа <http://www.seogadget.ru>.

8. You get signal - Get Trace Route Tool. Режим доступа <http://www.yougetsignal.com>.

9. WebMii. Find all information about any people and get their visibility score. Режим доступа <http://webmii.com>.

10. Pipl . The most comprehensive people search on the web. Режим доступа <https://pipl.com>.

11. Text.ru. Биржа копирайтинга. Режим доступа <http://text.ru/antiplagiat/>.

12. Text.ru. Описание сервиса. Режим доступа : <http://text.ru/seo>

***Анохіна І.Ю. Верифікація Інтернет - інформації.** Розглянуто питання верифікації інформації, одержуваної з Інтернету. Показано, якими способами і засобами можна перевірити сайти, визначити їх спрямованість, отримати інформацію про автора розміщеного матеріалу, оцінити ступінь плагіату в статті, книзі.*

***Anokhina I.Yu.** Verification of the Internet - information. The questions of verification of the information obtained from the Internet. It is shown in what ways and means you can check the sites, to determine it direction, to obtain information about the author of the material placed, to assess the degree of plagiarism in book.*

*Статья поступила в редакцию 20.09.2015  
Рекомендована к публикации д-ром техн. наук В.Н. Павлышом*