

ОБ ИДЕНТИФИКАТОРЕ ЦИФРОВОГО ОБЪЕКТА И ЕГО УСЛУГАХ

Ф. Ш. Аскеров, Р. Ш. Гасанова
Институт информационных технологий
Национальной академии наук Азербайджана, Баку

Представлена информация об идентификаторе цифрового объекта DOI, его имени, а также о Международном фонде DOI и регистрирующих его агентствах. Перечислены преимущества услуг Metadata retrieval, Content Registration, Similarity Check и Crossmark, предлагаемых исследователям некоммерческой организацией CrossRef. Приведена информация о сотрудничестве Института информационных технологий НАН Азербайджана с компанией CrossRef и статистическая информация о номерах DOI, полученных для научных трудов института.

Введение

Хотя обилие информации в современной интернет-среде создало для исследователей широкие возможности, здесь имеют место и определенные недостатки. Один из них – это отсутствие доступа к информации в онлайн-среде, связанное с удалением или изменением URL-адреса данной информации. Для предотвращения подобных проблем важным является внедрение идентификатора цифрового объекта – системы DOI (Digital Object Identifier). Кроме того, применение системы DOI позволяет исследователю быстро найти любую научную публикацию. Таким образом, выдав контенту имя DOI, можно обеспечить к нему стабильный доступ, независимо от того, был ли он удален или изменен его URL-адрес.

Система DOI была создана в издательской индустрии по инициативе трех бизнес-сообществ: International Publishers Association; International Association of Scientific, Technical and Medical Publishers; Association of American Publishers. Система была анонсирована в 1997 г. на Франкфуртской книжной ярмарке. В том же году был создан Международный фонд DOI (International DOI® Foundation, IDF) для управления и развития системы DOI (<https://www.doi.org/hb.html>).

В данной системе первые имена DOI для электронных версий статей начало выдавать регистрирующее агентство CrossRef в 2000 г. Тогда же DOI был утвержден организацией NISO (National Information Standards Organizations), а в 2010 г. – организацией ISO (International Organization for Standardization) как стандарт ISO 26324 и опубликован в 2012 г. Этот международный стандарт определяет синтаксис, описание и функциональные компоненты системы идентификаторов цифровых объектов. В нем также изложены общие принципы создания, регистрации и управления именами DOI.

1. Регистрационные агентства DOI

Имена DOI выдаются не Международным фондом DOI (IDF), а непосредственно его регистрационными агентствами. В настоящее время у фонда десять регистрационных агентств: Airiti Incorporation, CrossRef, China National Knowledge Infrastructure, DataCite, Entertainment Identifier Registry, Institute of Scientific and Technical Information of China, Japan Link Center и др. (https://www.doi.org/registration_agencies.html).

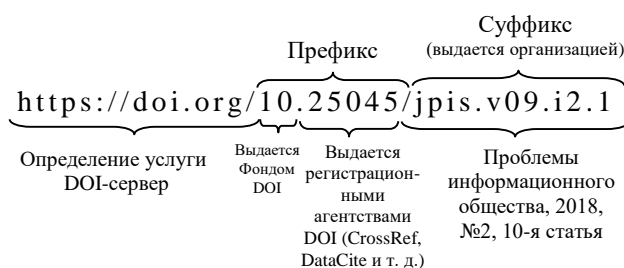
Агентство Airiti Incorporation, основанное в 2000 г., начало предоставлять услуги DOI в 2011 г., считается национальным и выдает имена DOI научным публикациям из Тайваня (<http://doi.airiti.com/en/index.html>).

Основанная в США в 2000 г. и управляемая компанией Publishers International Linking Association Inc. (PILA) некоммерческая организация CrossRef выдает имена DOI всем типам научных публикаций из любой страны. В настоящее время эта организация насчитывает более 14 тыс. членов из 140 стран (<https://www.crossref.org/about/>). Регистрационное агентство DataCite выдает имена DOI ссылкам и научным данным, а Entertainment Identifier Registry – коммерческим фильмам и видео (<https://datacite.org/>).

2. DOI – идентификатор цифрового объекта

Идентификатор цифрового объекта DOI – это номер, выданный интеллектуальному документу в Интернете и который значительно упрощает поиск и использование материалов в виртуальной среде. Эта технология в основном используется в научной среде и включает ссылку на конкретный объект. Объект, обладающий идентификатором, может быть удален только издателем, т. е. организацией, разместившей его (https://www.doi.org/overview/DOI_article_ELIS3.pdf).

Номера DOI состоят из двух частей – префиксов и суффиксов. 10 цифр, перечисленных в гиперссылке после доменного имени «doi.org», указывают на имя DOI, а следующая часть – на регистрационное агентство, от которого получено имя, далее указываются организация и интеллектуальный продукт (рисунок).



Структура идентификатора цифрового объекта DOI

3. Услуги, предоставляемые CrossRef

Некоммерческая организация CrossRef, которая является одним из агентств по регистрации DOI, предоставляет исследователям следующие услуги: Metadata retrieval, Content Registration, Reference linking, Cited-by, Funder Registry, Similarity Check, Crossmark и Event Data (<https://www.crossref.org/services/>).

Metada retrieval – сервис для получения метаданных. Метаданные членов CrossRef доступны всем пользователям с помощью различных инструментов, эти данные можно искать и повторно использовать.

Члены CrossRef регистрируют здесь свой контент, поэтому о его существовании знает весь мир. Так называемые метаданные, отправленные участниками, собираются и хранятся здесь стандартным способом. Метаданные содержат имя DOI для каждого контента, поэтому DOI ссылается на него, даже если контент перемещается на другой сайт. Эти метаданные доступны с помощью специальных приложений, которые, в свою очередь, позволяют людям и машинам (компьютерам) включать их в исследовательские инструменты и сервисы. Метаданные не содержат полный текст содержания.

Чтобы использовать службу сбора метаданных, необязательно нужно быть членом CrossRef. Приложение REST API этой службы помогает другим системам получать более полные и обновленные метаданные, принадлежащие издателям. Такие метаданные используются поисковыми службами, системами отслеживания рукописей, про-

граммным обеспечением для управления библиографией, инструментами профиля автора, специализированными базами данных и сетями научного обмена. Это помогает пользователям находить интересующие их исследования, ссылаться на них и оценивать. Таким образом, издатели делают свой контент доступным для всех и предоставляют его для использования различными службами, просто отправляя метаданные на один адрес. Служба Metada retrieval помогает исследователям поддерживать, разрабатывать, публиковать и делиться своей работой.

Многие инструменты и службы используют эти метаданные для поиска, сводки, совместного использования и анализа. Наиболее широко используемые варианты приложения REST API следующие: интеллектуальный анализ текстов и данных, помощь посредством спонсоров, идентификация авторских публикаций.

Интеллектуальный анализ текстов и данных (Text and Data Mining, TDM) – это сбор и анализ данных из большого количества документов. Использование DOI в качестве основы такого анализа дает исследователю несколько преимуществ:

- удаление дубликатов; исключает повторяющиеся документы, найденные на различных сайтах, независимо от их формата представления (PDF, HTML и др.);
- постоянная доступность информации; использование DOI в качестве ключа позволит исследователю найти и проверить источники контентов, охватываемые TDM, даже через несколько лет, если традиционные URL-адреса не работают;
- нет необходимости работать с самими документами для документирования, обмена и сравнения;
- механизм обеспечения продуктивности результатов TDM с использованием исходных документов;
- механизм мониторинга влияния обновлений и изменений.

Исследователи больше заинтересованы в применении TDM к научному контенту. Для этого требуется автоматический доступ к полному тексту множества статей. Формат полного текста зависит от участников. Метаданные помогают исследователю получить доступ к этому контенту и позволяют участникам предоставлять его.

Согласно статистике, по состоянию на октябрь 2019 г. более 642 млн чел. подали заявки на сервис поиска по метаданным.

Content Registration – услуга регистрации контента. Член CrossRef может сделать свой контент доступным в Интернете, заключать множество двусторонних соглашений с другими организациями. Сервис также позволяет членам организации регистрировать и обновлять свои метаданные с помощью различных интерфейсов. Здесь можно зарегистрировать книги, материалы конференций, базы данных, диссертации, грантовые проекты, журналы и журнальные статьи, рецензии, отчеты, препринты, стандарты, публикации, находящиеся в ожидании, и т. д. Член организации, регистрирующий имена DOI, вводит в систему следующую информацию: даты, заглавия, авторов, организации, спонсоров и местоположение в Интернете.

По состоянию на октябрь 2019 г. количество зарегистрированного контента составляет более 109 млн. Из них 72 149 – это количество журналов, 79 777 806 – журнальные статьи, 1 380 292 – книги, 68 280 – материалы конференций, 6 048 716 – тезисы конференций, 296 162 – диссертации (<https://doi.org/10.13003/y8yugwm5>).

Similarity Check – услуга проверки на подобие. Данная услуга, предлагаемая CrossRef и основанная на программном продукте iThenticate компании Turnitin, представляет редакторам удобный инструмент для обнаружения плагиата.

Исследователь должен быть уверен в оригинальности прочитанного. Услуга Similarity Check связывает издательства и редакторов с инструментом iThenticate. Исследуемый контент проверяется с помощью базы данных, содержащей миллионы опубли-

ликованных научных статей и миллиарды веб-страниц. Любое сходство отражается в отчете, что позволяет редакторам перед публикацией проверить оригинальность работы. Similarity Check для издательств и авторов обеспечивает сохранение конфиденциальности, связанной с плагиатом, а также ясность для читателей с точки зрения обеспечения подлинности работы.

Количество членов CrossRef, которые пользуются этим сервисом, на 2021 г. составляет 1532 (<https://doi.org/10.13003/y8ygw5>). Сервис предоставляет читателям быстрый и легкий доступ к текущему статусу контента, включая любые изменения, сокращения и обновления. Как известно, исследования с течением времени развиваются. Даже после публикации статьи могут быть обновлены из-за новых данных, а также на основании каких-либо изменений или полученных отрицательных отзывов. Используя эту услугу, исследователи и библиотекари могут легко отслеживать изменения в содержании, которое они читают, включая клинические испытания, ORCID iD и многое другое, а также узнать, кто финансирует исследование и какая лицензия применяется к содержанию (<https://www.crossref.org/services/>).

Эта и другая подобная дополнительная информация может быть включена в уже опубликованный контент позже. Данную услугу важно использовать для того, чтобы быть уверенным в надежности исследования. Издатели вводят все эти метаданные в систему. Такую информацию можно получить в онлайн-формате или в формате PDF, нажав кнопку Crossmark, которая используется во всех издательствах и платформах. Эта кнопка выдает читателю информацию об изменениях и гарантирует издателю постоянное управление своим контентом.

Данная информация находится в открытом доступе, поэтому ее могут использовать и другие системы. Исследователи также могут использовать Crossmark для интерпретации работы, полностью доверяя ей. Здесь можно следить за важными обновлениями контента и легко получать доступ к добавленной информации.

Заключение

Исследователи, издатели и редакторы могут воспользоваться услугами перечисленных выше некоммерческих организаций CrossRef. Возможность проверки на плагиат научных работ, отслеживание исправлений и изменений позволяют исследователю оценить интересующую его область анализа. Имя DOI также содержит цифровой идентификатор автора, такой, как ORCID iD, что упрощает связи исследователей в научных социальных сетях. Получив фиксированный идентификатор для публикаций, можно получить доступ к полнотекстовому контенту и выполнить интеллектуальный анализ текстов и данных.

Следует отметить, что Институт информационных технологий НАН Азербайджана оказывает услуги DOI с 2017 г., подписав контракт с CrossRef. Каждая статья в журналах «Проблемы информационных технологий» и «Проблемы информационного общества», публикуемых в институте, получает номер DOI. Общее количество полученных номеров DOI – 543, включая номера DOI для материалов конференций, проведенных институтом, и книг. Таким образом, 274 из этих номеров DOI были получены для статей, 239 – для материалов конференций и 30 – для книг.