

2020



IDN-домены: зачем они нужны и как работают

Материал подготовлен **Координационным центром доменов .RU/.РФ**
при поддержке проекта **Поддерживаю.РФ**



Что такое доменное имя

Доменные имена – удобный для людей способ идентифицировать различные ресурсы в интернете. IP-адрес, изначально созданный для указания конкретного компьютера или ресурса в сети, состоит из набора чисел и поэтому труден для запоминания человеком.

Доменное имя выглядит как последовательность текстовых «меток», чаще всего понятных слов или сокращений, разделенных точками, например, www.test.ru. Каждая такая «метка» (www, test, ru) представляет один из уровней в иерархии доменных имен, и такие уровни располагаются справа налево. В самом простом случае доменное имя состоит из двух частей – например, test.ru. Правая часть ru называется доменом первого или верхнего уровня, доменной зоной, TLD (Top Level Domain), а левая часть содержит уникальное имя ресурса, и в нашем примере – это домен второго уровня test.

Интересный факт

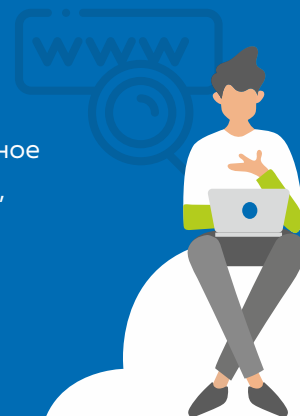
На самом деле, существует еще и **домен нулевого уровня**, также называемый корневым доменом (root domain). Однако в соответствии со стандартами, его имя представляет собой пустую строку, и сам корневой домен указывается как точка в начале доменного имени (на примере «test.ru.» – это первая точка справа). Но сейчас все программное обеспечение позволяет не вводить такую точку, в большинстве случаев добавляя ее к доменному имени автоматически.



Также могут существовать более сложные доменные имена. Например, доменное имя www.corp.mail.ru содержит домен верхнего уровня ru, домен второго уровня mail (которые вместе составляют бренд бесплатной почты mail.ru), а также домен третьего уровня corp, относящийся не ко всей бесплатной почтовой службе, а только к самой компании. В данном случае домен четвертого уровня www – и это адрес ресурса, на котором находится веб-сервер – программа, обеспечивающая доставку разнообразного контента, в первую очередь веб-страниц, по протоколам HTTP/HTTPS.

Это интересно:

www (World Wide Web) – типовое название веб-сервера, придуманное в 1989 году Тимом Бернерсом-Ли, и вместе с идентификаторами URI, протоколом HTTP и языком HTML образующее основы всемирной паутины. Поскольку со временем Web стал основным сервисом интернета, добавлять **www** перестали. Сегодня при обращении к доменному имени без **www** трафик автоматически направляется на соответствующий веб-сервер по протоколам HTTP/HTTPS.



Обеспечивает преобразование доменных имен в IP-адреса и обратно **«система доменных имен» (DNS – Domain Name System)**. Специальные серверы DNS¹ расположены в ключевых точках сети Интернет и отвечают на запросы о том, какой IP-адрес имеет ресурс с заданным именем. Этот процесс называется трансляцией доменных имен.

Разработку правил регистрации и использования доменных имен в каждой доменной зоне, а также хранение базы зарегистрированных доменных имен обеспечивают специальные организации – **регистратуры соответствующего домена верхнего уровня**. Например, регистратурой национальных доменов .RU и .РФ является Координационный центр доменов .RU/.РФ.

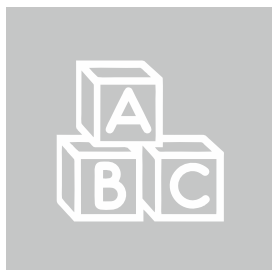
Конечным пользователям услуги по регистрации доменных имен оказывают компании-регистраторы, которые проходят процедуру аккредитации по специальным правилам регистратуры каждого домена верхнего уровня (TLD).

Общую координацию по созданию и поддержке TLD осуществляет **Корпорация по управлению доменными именами и IP-адресами (ICANN – Internet Corporation for Assigned Names and Numbers)**. Она же обеспечивает поддержку «корневой зоны» – базы всех существующих TLD и связанных с ними DNS-серверов.



¹ <https://ru.wikipedia.org/wiki/DNS>

Домены на национальных языках (IDN)



Исторически интернет был разработан в США, поэтому изначально в доменных именах предполагалось использовать только символы латинского алфавита **A-Z**, дефис и цифры, входящие в набор **символов ASCII**² (American Standard Code for Information Interchange). Однако со временем ситуация изменилась.

Интернет постепенно охватил всю планету, и им стали пользоваться миллионы обычных людей, в том числе те, кто не владеет английским языком. Содержание веб-сайтов, тексты почтовых сообщений, интерфейсы различных онлайн-сервисов со временем стали доступны пользователям **на разных языках**, чтобы каждый смог выбрать наиболее привычный и удобный для себя. Интернет-пользователи отдают предпочтение своим национальным языкам, в том числе и в интернет-идентификаторах: доменных именах сайтов, адресах электронной почты, учетных записях, и это стало точкой отсчета для появления доменов на национальных языках.

Эксперименты по решению этой проблемы шли с 1998 года, и уже в 2004 году появились доменные имена второго уровня, позволяющие использовать расширенный латинский алфавит, включающий, например, важную для испаноговорящих пользователей букву **ñ**. При этом домены верхнего уровня оставались латинскими. Такие домены называют **частичными IDN** или **IDN SLD**.



Следующим шагом стало появление **полных IDN-доменов** (IDN - Internationalized Domain Names), в которых уже любая часть доменного имени, начиная от названия доменной зоны и заканчивая доменом 3-го, 4-го и ниже уровня, может быть интернационализированной, то есть содержать символы национальных алфавитов. В 2010 году были одобрены национальные (страновые) IDN на русском и арабском языках. Появились домены верхнего уровня: **امارات**. (ОАЭ), **السعودية**. (Саудовская Аравия), **مصر**. (Египет).

²<https://ru.wikipedia.org/wiki/ASCII>

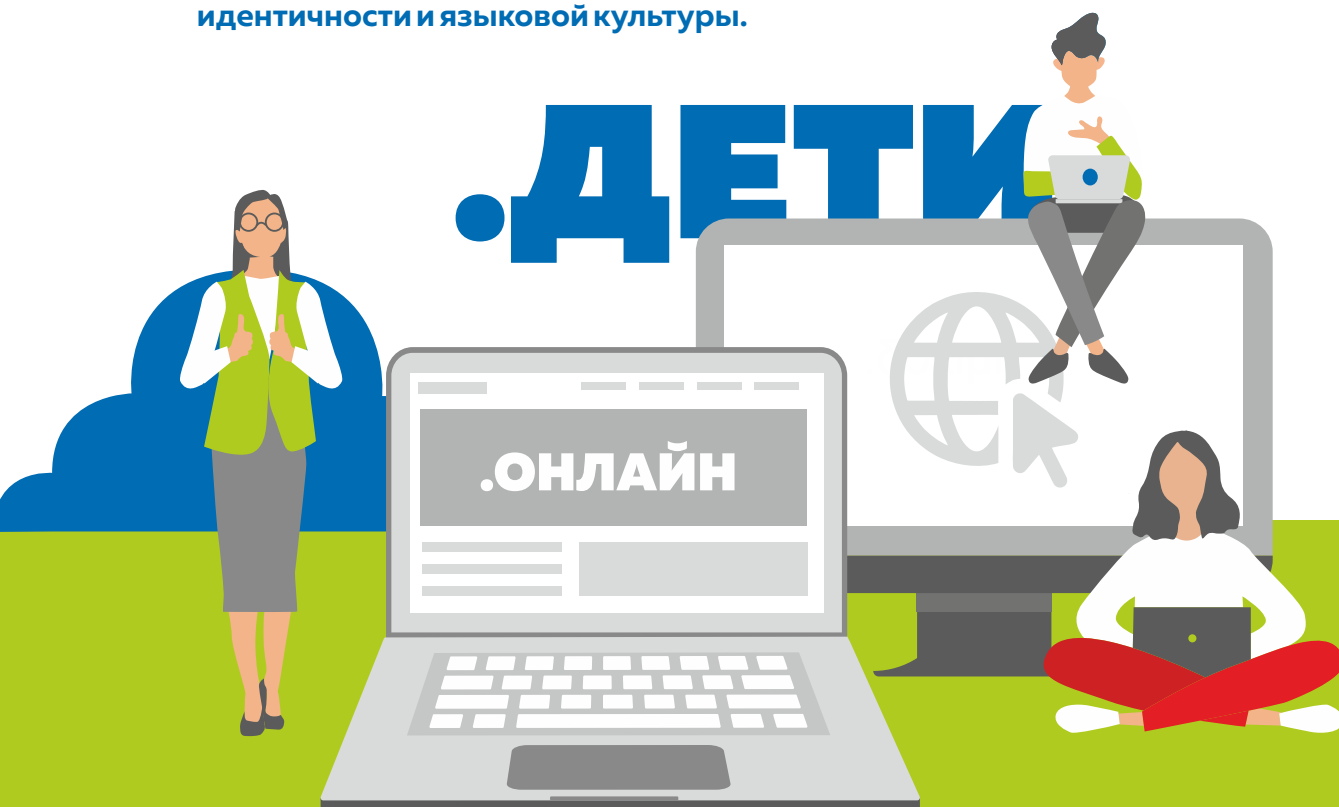


Первым кириллическим доменом верхнего уровня стал российский **.РФ**, открытый 12 мая 2010 года. В этом году ему исполнилось десять лет!

Вслед за национальными появились и первые IDN gTLD – доменные зоны общего назначения, не относящиеся к конкретным странам, но использующие национальные алфавиты. Например, в 2014 году стали доступны кириллические домены **.ДЕТИ**, **.МОСКВА**, **.САЙТ**, **.ОНЛАЙН** и другие.

Пользоваться IDN-доменами значительно удобнее, чем латинскими. Например, доменное имя в **.РФ**, где разрешены к использованию символы русского алфавита, прочесть, запомнить и затем без ошибок набрать в адресной строке браузера значительно проще. Не возникает и проблем с использованием множества русских слов, содержащих **ч, щ, ш, я, ю, э** и другие буквы, которые могут транслитерироваться (переводиться) в латинский шрифт несколькими способами (например, слово матрешка как *matreshka* или *matryoshka*, а алфавит как *alphavit* или *alfavit*).

В более сложных языках, таких как, например, малайский, арабский (включая написание справа налево) или использующий иероглифическое письмо китайский, IDN-домены дают возможность миллиардам людей, не знающим английского языка, **использовать интернет в рамках своей национальной идентичности и языковой культуры**.



Как работают IDN-домены

В основе обработки компьютером любого текста, в том числе доменных имен, лежит **таблица символов**. С помощью этой таблицы программы, установленные на компьютере, соотносят изображение символа с его цифровым кодом, который хранится и обрабатывается ими. Перед выводом каждого символа на экран или печать цифровой код с использованием той же таблицы и набора шрифтов преобразуется в изображение, которое появляется на экране или бумаге.



Лежащая в основе системы доменных имен **таблица ASCII** – одна из первых таблиц символов, разработанная в 60-годах прошлого века. Она включает символы латинского (английского) алфавита, цифры, знаки пунктуации и технические символы – всего 128 символов. Это ограничение связано с тем, что интернет появился в США, а код ASCII-символа изначально должен был занимать один байт, причем это был не современный 8-битный, а 7-битный код, использовавшийся в компьютерах 70-х годов прошлого века.



IDN-домены используют более новую таблицу символов **Unicode**³, разработанную в 1991 году. Эта таблица содержит 65 536 символов, включая знаки почти всех письменных языков мира, а также математические символы, символы музыкальной нотации и многие другие. Часть кодов зарезервирована для использования в будущем.

Коды в стандарте Unicode разделены на несколько областей. Область с кодами от U+0000 до U+007F содержит символы набора ASCII, и коды этих символов совпадают с их кодами в ASCII. Далее расположены области символов других систем письменности, знаки пунктуации и технические символы.

³ <https://ru.wikipedia.org/wiki/Юникод>

Предупрежден — значит вооружен

Появление возможности использовать доменные имена с символами национальных алфавитов, в том числе символами нескольких алфавитов в одном доменном имени, создало и несколько проблем. Все они так или иначе относятся к безопасности, а точнее, к возможности создавать визуально похожие, но отличающиеся в кодах символов доменные имена и с их помощью заманивать пользователей на фальшивые сайты, где те могут ввести данные своих банковских карт или раскрыть мошенникам иную важную информацию.

Для этого используются, например, **омоглифы** — символы, которые из-за сходства по размеру и форме кажутся идентичными или похожими друг на друга до степени смешения. Они часто встречаются при смешивании латинского, кириллического и греческого алфавитов. Самым простым примером омоглифа являются русские и латинские буквы а, о, е, с, которые в обоих алфавитах выглядят одинаково.

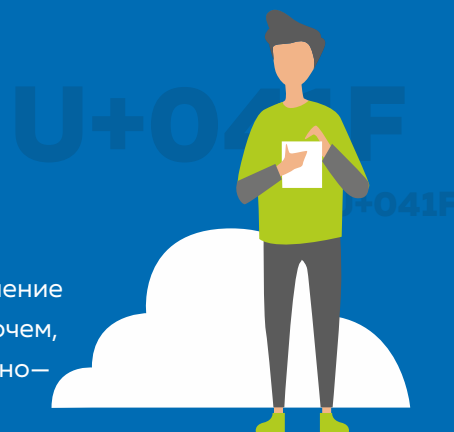
Чтобы предотвратить создание доменных имен с омоглифами, регистратуры доменов верхнего уровня ограничивают список символов одним или несколькими совместимыми алфавитами. У каждой регистратуры есть таблица допустимых символов (IDN Table), а также правила регистрации доменных имен в конкретной доменной зоне, в соответствии с которыми разрешена регистрация доменных имен, и, например, в домене .РФ разрешено использовать только русские буквы, арабские цифры и дефис. Полный список таблиц допустимых символов можно найти на сайте **IANA**⁴ (Internet Assigned Numbers Authority).

⁴ <https://www.iana.org/domains/idn-tables>



Это интересно:

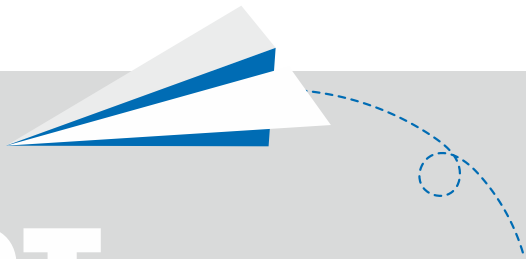
Unicode постоянно развивается, и в него добавляются новые символы. Хотя формы записи UTF-8 и UTF-32 позволяют кодировать до 2^{31} (**2 147 483 648**) символов, было принято решение использовать лишь 1 112 064 для совместимости с UTF-16. Впрочем, на данный момент даже и этого количества более чем достаточно— в версии 13.0 используется всего **143 859** кодовых позиций.



Поскольку в системе DNS ранее использовалась только таблица ASCII, была разработана дополнительная система перекодировки из Unicode в ASCII и обратно, названная **Punycode**. Ее задача в том, чтобы программы и интернет-сервисы, которые не были изначально готовы к работе с IDN-доменами, могли понимать имена, использующие Unicode. Такое требование называется **«обратная совместимость»** и применяется к большинству нововведений в интернете, так как инфраструктура сети крайне обширна, не контролируется из единого центра, и нет никакой гарантии, что на всех серверах или компьютерах пользователей установлено программное обеспечение, поддерживающее современные стандарты.

Получающиеся в результате преобразования в ASCII с помощью алгоритма Punycode строки называются **A-метками (A-labels)**. Характерной особенностью A-метки является то, что она всегда начинается со следующих четырех символов «xn--». Эти символы называются **префиксом ACE**. Punycode — обратимое преобразование: можно преобразовать строку Unicode в A-метку и снова преобразовать ее в строку Unicode (которая называется **U-меткой**).

Пример кодировки:



Привет

U+041F U+0440 U+0438 U+0432 U+0435 U+0442

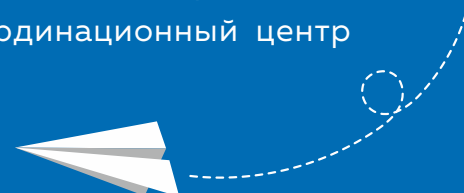
В алфавитах некоторых языков отдельные фонемы (звуки) записываются в виде двух символов, называемых диграфом. Если эти диграфы имеют одинаковое толкование на всех языках, использующих данный алфавит, специальные алгоритмы обработки (нормализации) Unicode устраняют различия и обеспечивают их согласование. Однако существуют языки, например, персидский, малайский или тамильский, в которых одна и та же фонема может записываться несколькими способами – например, одним диграфом и сочетанием двух символов. Причем отображается и диграф, и набор символов одинаково, отличия проявляются лишь в кодах символов.

Универсальное принятие

DNS-серверы обрабатывают запросы только в ASCII, а пользователи хотят видеть IDN-имена на знакомых им национальных языках. В результате большая часть программного обеспечения, связанного с интернетом, должна показывать и позволять вводить IDN-домены в виде **U-меток** (например, [тест.рф](#)), но хранит и обрабатывает их в виде **A-меток** (например, xn--e1aybc.xn--p1ai). При этом до сих пор осталось немало программ и сервисов, которые не полностью поддерживают работу с IDN-именами: не разрешают вводить, не умеют обрабатывать, неверно отображают, выдают сообщения об ошибке или неверно введенном доменном имени. Аналогичные проблемы возникают и с новыми доменами общего пользования (new gTLD) на латинице.

Состояние, при котором все программы и сервисы полностью поддерживают IDN-домены и Unicode, в том числе в почтовых адресах, а также новые латинские домены верхнего уровня (New gTLD), называется **«универсальным принятием» (Universal Acceptance)**. Специалисты в России и на международном уровне работают над тем, чтобы IDN-домены и Unicode стали полностью поддерживаться всем программным обеспечением, всеми приложениями и сервисами, использующими интернет.

Координационная группа по универсальному принятию (Universal Acceptance Steering Group), созданная в 2015 году по инициативе интернет-сообщества и включающая представителей ведущих мировых IT-компаний, таких как Microsoft, Google и Apple, координирует доработку программного обеспечения, направленную, в том числе, на повсеместную поддержку IDN. В нашей стране помощь в доработке программ и интернет-сервисов для работы с русскоязычными доменами также занимается Координационный центр доменов .RU/.РФ в рамках проекта [Поддерживаю.рф](#).



Как зарегистрировать доменное имя

Шаг 1

Придумать доменное имя и выяснить, свободно ли оно.

В этом вам поможет сервис WHOIS на сайте регистратуры домена верхнего уровня.

Шаг 2

Выбрать компанию-регистратора.

Для этого нужно изучить условия оказания услуг и цены у разных аккредитованных регистраторов, почитать о них отзывы в интернете. Актуальный список аккредитованных регистраторов публикуется на сайте регистратуры домена верхнего уровня.

Шаг 3

Зарегистрировать доменное имя.

Для регистрации доменного имени вы подаете заявку на сайте выбранного регистратора, следуя установленной им процедуре. Имейте в виду: все сведения, которые вы указываете, должны быть достоверны – недостоверные сведения могут стать причиной отказа в регистрации или же, в дальнейшем, аннулирования регистрации доменного имени.

После регистрации доменного имени вы станете его администратором и сможете им управлять. Только вы будете решать, как использовать доменное имя и созданный на нем интернет-ресурс, и кто будет осуществлять его техподдержку.

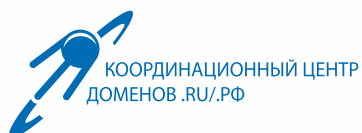
Г
Е
В
О
С



Материал подготовлен **Координационным центром доменов .RU/.РФ**
при поддержке проекта **Поддерживаю.РФ**

тел. +7 495 7302971

cctld.ru | кц.рф
поддерживаю.рф



Поддерживаю.РФ