07/2019

nic//report

Ausgabe 5 // **RESEARCH & DEVELOPMENT**

nic.at

nic // report – Inhalt

R&D Leiter Alexander Mayrhofer im Interview 02
Facts & Figures 04
Schwerpunkte von R&D 06
Domainvorschlagstool 08

Sehr geehrte Damen und Herren!

Sie fragen sich, wo die vielen innovativen Forschungsprojekte, Ideen und Zahlen rund um unser Unternehmen ihren Ursprung haben? Welche Tools unsere Prognosen und strategischen Entscheidungen beeinflussen? Wir laden Sie mit diesem nic//report ein, einen Blick hinter die Kulissen der Research & Development Abteilung zu werfen und zu erfahren, mit welchen Themen wir uns dort auseinandersetzen. Abteilungsleiter Alexander Mayrhofer verrät uns außerdem, wie neue Ideen überhaupt entstehen und in weiterer Folge zu Projekten werden, wie man selbst innovativ bleibt und was uns in Zukunft erwartet.

Auch wir staunen immer wieder, was in der R&D Abteilung „gebastelt" und kreiert wird – hätten Sie gedacht, dass die hübschen Pusteblumen oberhalb den Zusammenhang von einzelnen, in Domains verwendeten Wörtern darstellen? Hier erhalten Sie also nicht nur umfassend Einblick in die R&D Abteilung, sondern können möglicherweise auch ihr Know-How erweitern. Denn: Wer mit Begriffen wie Blockchain, RDAP oder DNS over HTTPS gekonnt jonglieren kann und über aktuelle Trends Bescheid weiß, erntet beim nächsten Business Termin oder Abendessen mit Freunden mit Sicherheit den einen oder anderen anerkennenden Blick.

In diesem Sinne: Viel Spaß beim Lesen der folgenden Seiten!

Richard Wein und Robert Schischka
(Geschäftsführer **nic.at**)

R&D: EIN BLICK HINTER DIE KULISSEN
Alexander Mayrhofer im Interview

Zunächst ganz allgemein: Was sind die Aufgaben der R&D Abteilung bei **nic.at**?
Unsere Tätigkeiten lassen sich in drei Säulen unterteilen: Einerseits befassen wir uns mit der Entwicklung und der Verbesserung bestehender Produkte, wie unserer **tldbox** oder **RcodeZero DNS**. Die zweite Säule dreht sich um die Themen Community und Standardisierung: Auf internationaler Ebene gestalten wir bei CENTR mit, und sind im Rahmen der Internet Engineering Task Force (IETF) an der Weiterentwicklung des Internet beteiligt. Für die nationale Community veranstalten wir Events, etwa in Form der »DNSheads Vienna« meetups. Unsere dritte Säule betrifft die »Forschungsarbeit« – also z. B. neue Erkenntnisse aus unseren Daten zu gewinnen, aber auch aktuelle Themen wie Identity Management.

Unsere Aufgabe ist es, sehr komplexe Dinge so aufzubereiten, dass wir die Erkenntnisse daraus sinnvoll intern nutzen bzw. an unsere Registrare und Partner weitergeben können.

An welchen Projekten arbeitest du im Moment?
Derzeit beschäftige ich mich intensiv mit der sogenannten »DNS Magnitude«, eine Art Maßzahl für die Beliebtheit von Domains. Mithilfe eines speziellen Algorithmus wird dabei jede Domain auf einer Skala zwischen 0 und 10 eingeordnet – eine einfache Möglichkeit zu beurteilen, wie intensiv eine Domain genutzt wird. In weiterer Folge kann das beispielsweise auch für Upselling durch Registrare genutzt werden – eine populäre Domain sollte mit Produkten wie »**RcodeZero DNS**« natürlich entsprechend abgesichert werden. Schließlich wäre es z. B. bei einem gut besuchten Online-Shop für Kunde und Betreiber ärgerlich, wenn die Domain vorübergehend nicht aufrufbar wäre.

Ein zweites, großes Thema ist »verschlüsseltes DNS«: Fast jede Aktivität im Internet startet mit DNS-Abfragen, und diese Informationen gehen dabei meist im Klartext durch das Netz. Mit »DNS over TLS« und »DNS over HTTPS« gibt es dabei gleich zwei neue Protokolle, die das ändern wollen. Vor allem DNS over HTTPS hat dabei aber das Potenzial, dass in Zukunft der Gutteil der DNS-Abfragen nur mehr von wenigen großen Betreiber beantwortet wird – eine Entwicklung, die von vielen zumindest mit Skepsis betrachtet wird.

Woher kommen die Ideen, die später zu Projekten werden?
Manche entstehen aus dem Tagesgeschäft oder einfach dann, wenn man etwas Zeit hat, um sich mit Neuem zu beschäftigen. Auch der Austausch mit Gleichgesinnten, die ähnliche Probleme lösen möchten, also Ideen aus der Community, sind oft sehr inspirierend. Innovation bedeutet aber nicht immer, etwas komplett Neues zu erfinden, meist werden bestehende Dinge einfach neu kombiniert. Ein Beispiel dafür ist unser Data Warehouse – wir beginnen jetzt damit, unsere internen Erkenntnisse neu zu kombinieren, und sie auch Registraren und der Öffentlichkeit anzubieten.

Wie viele dieser Ideen werden tatsächlich umgesetzt?
Da kann man das Konzept einer Innovationspyramide mit einem Faktor 10 heranziehen: Aus 3.000 Ideen werden 300 genauer untersuchte Themen, daraus 30 konkrete, größere Projekte, die dann in 3 völlig neue strategische Ausrichtungen für ein Unternehmen münden können. Es geht also viel um's Ausprobieren und darum, Ideen auch wieder zu verwerfen und wieder etwas anderes zu versuchen.

Was ist deine Motivation, neugierig zu bleiben? Wie bleibt man innovativ?
Ich mag Dinge, die zunächst sehr komplex sind und die ich so vereinfachen kann, dass sie sinnvoll genutzt werden können, und – das klingt jetzt kitschig – einfach »schön« sind. Unsere Motivation ist, etwas Nützliches zu produzieren, das am Ende tatsächlich eingesetzt werden kann und zum Erfolg des Unternehmens beiträgt. Und innovativ, denke ich, bleibt man dann, wenn man sich täglich zigmal Fragen stellt wie »Kann ich das verbessern? Löse ich ein Problem? Was ist der Benefit – wie profitieren Nutzer, Registrare, unser Unternehmen?« Was mich freut, ist, dass Google aktuell ein Protokoll-Detail bei verschlüsseltem DNS implementiert, das ich standardisiert habe.

Ist die Gefahr nicht groß sich zu verzetteln?
Ein klares »Ja«. Man darf nicht zu viele Themen gleichzeitig angehen – die Verlockung ist natürlich groß, weil so viel neu und spannend ist. Innovation hat immer auch mit dem Risiko von »stranded work« zu tun – mutig über den Tellerrand blicken, aber Ideen und Projekte auch wieder aufgeben können, wenn sie keinen Sinn stiften.

Kannst du für uns einen Blick in die Zukunft werfen?
Ich kann's zumindest versuchen. Machine Learning wird bald auch im normalen Geschäftsbereich eine alltägliche Rolle spielen, auch wir nutzen es schon in unserer Prognose. Die Herausforderung wird dabei sein, Entscheidungen, die von Maschinen getroffen werden, als Mensch zu verantworten – wenn mir ein Algorithmus jemals untersagt, ein Zugticket zu kaufen, dann will ich eine sehr fundierte Begründung hören.

Außerdem wird Identity Management ein wichtiges Thema werden. Die Öffentlichkeit wird sensibler, was das Thema der Datensicherheit und Datensouveränität anbelangt – da wird sich in naher Zukunft sehr viel tun.

ALEXANDER MAYRHOFER
leitet die R&D Abteilung seit 2002 und ist unter anderem auch in der IETF (Internet Engineering Task Force) aktiv. Er ist als Speaker bei zahlreichen Veranstaltungen unterwegs und begeistert dort nicht nur mit seinen Vorträgen, sondern immer auch mit seinen anschaulichen Illustrationen.

*Legendär Alexanders »Domain-Lagune«: Sie veranschaulicht die unendlichen Weiten des **.at**-Namespaces - und, dass die tatsächlich registrierte Anzahl von Domains kaum daran herankommt. Als »Insel der Seligen« die 1,3 Million registrierten **.at**-Domains, die vom Atoll der gelöschten Domains und der Lagune der Whois- und EPP-Abfragen zu freien Domains gehören. Rundherum das DNS-Meer von 300 Millionen NX-Domains pro Monat. Das sind Domainnamen, die auf unseren DNS-Servern abgefragt werden, aber nicht existieren. Draußen auf hoher See lauern außer Meeresungeheuer noch Unmengen an noch nicht registrierten und nie abgefragten Domains. Man sieht also, dass das Potenzial an **.at**-Domains bei weitem noch nicht ausgeschöpft ist.*

FACTS & FIGURES AUS DER R&D ABTEILUNG

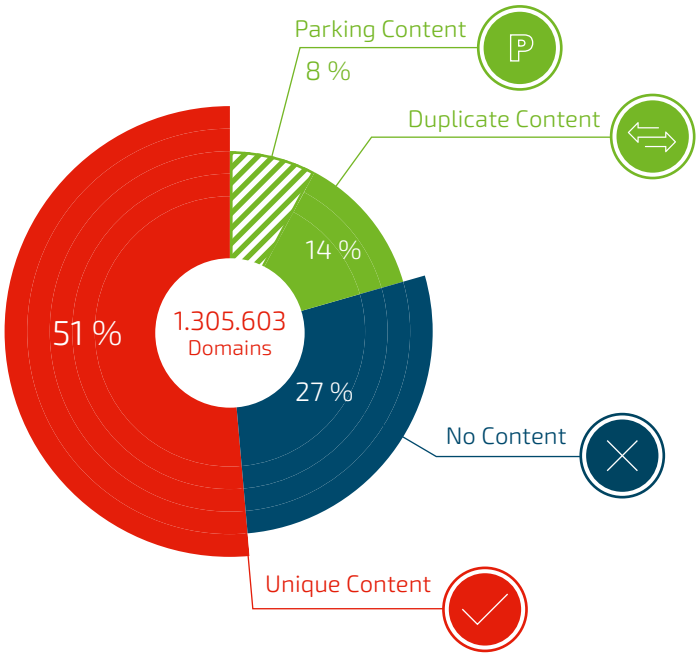
Top 5 Vornamen in .at-Domains vs. Österreichs beliebteste Vornamen

	Domains	Vornamen*
1	max	Michael
2	peter	Lukas
3	martin	Daniel
4	mark	Thomas
5	michael	Florian

1	maria	Julia
2	eva	Anna
3	anna	Katharina
4	andrea	Lisa
5	elisabeth	Sarah

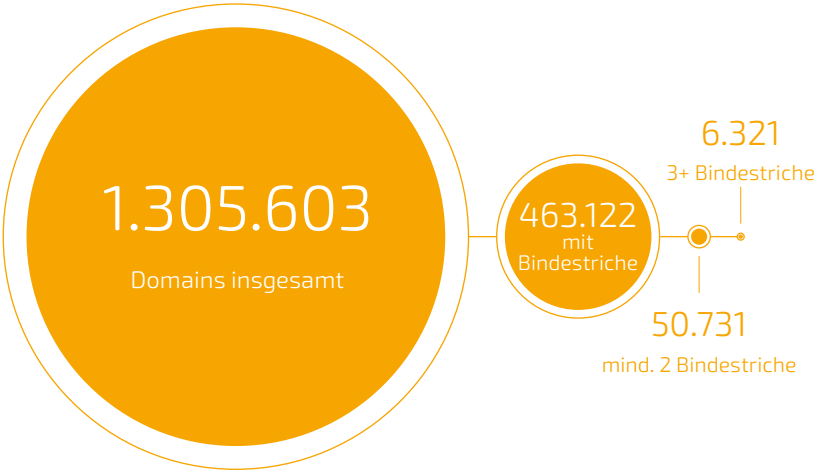
* Quelle: Statistik Austria

Websites hinter den .at-Domains



Domains mit Bindestrichen

(Stichtag 25.6.2019)

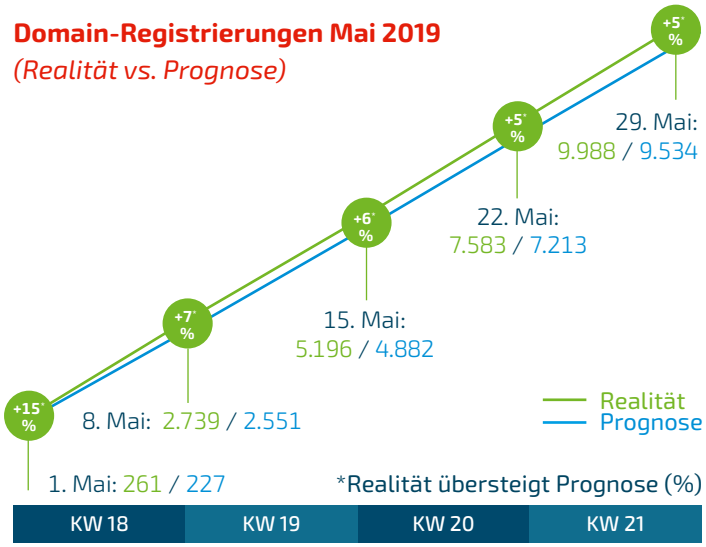


DNS Magnitude pro Domain

(beliebteste Websites der .at-Zone auf einer Skala von 0 bis 10)

	Domain	Magnitude
1	google.at	9,1
2	pinterest.at	9,0
3	airbnb.at	8,8
4	ebay.at	8,5
5	expedia.at	8,5
6	tripadvisor.at	8,5
7	orf.at	8,3
8	groupon.at	8,2
9	derstandard.at	8,1
10	wienerzeitung.at	8,1

Domain-Registrierungen Mai 2019
(Realität vs. Prognose)



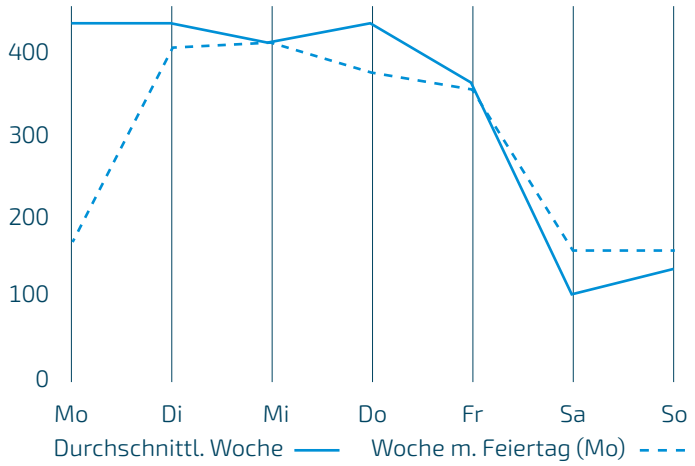
Top 30 Begriffe in .at-Domains

Anzahl der möglichen .at-Domains*

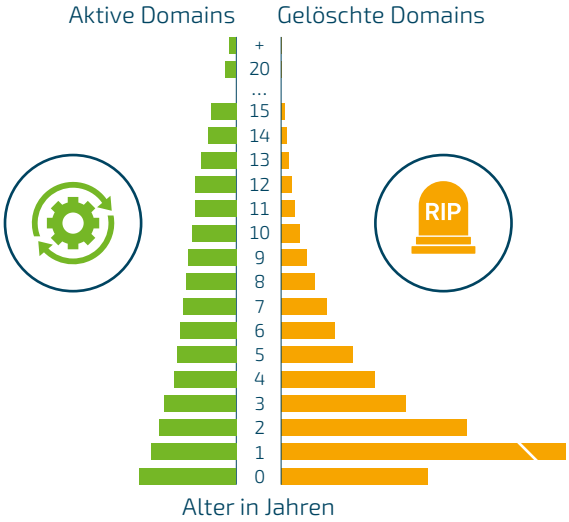
111.444.219.848.545.291.112.918.149.658.401.217.019.177.846.881.717.006.276.548.100.629.318.214.534.968.256.903.948.922.840.416.256

*mindestens

Domain-Registrierungen im Wochenverlauf
(Durchschnitt und mit Feiertag am Montag)



Demografie der .at-Zone



Domains mit gültigem Zertifikat

53,9 % von 1.306.503 Domains *(Stichtag 12.6.2019)*

Durchschnittliche Schuhgröße
(im R&D- Team)

AKTUELLE R&D PROJEKTE:
Schwerpunkte der Mitarbeiter

PHILIPP ADAM
unterstützt das Team seit 2019 im Bereich Data Warehouse als »Junior Research Engineer« und macht aktuell an der WU Wien seinen Master im Bereich »Information Systems«.

Wissenswertes zu Domains

Wie werden **.at**-Domains genutzt? Wie schnell landet sogenannter »unique content« nach der Domain-Registrierung auf der Seite? Fragen wie diese beantwortet unser **.at**-Webcrawler, der sich die Startseite der Website hinter jeder Domain holt und deren Inhalt analysiert. Auf die Lebensdauer einer Domain hat z. B. Einfluss, ob hinter dieser eine Website mit Content steht. Diese Informationen sind auch für Registrare und Partner interessant, um Einblick in die eigenen Portfolios, aber auch potentielle Upselling-Möglichkeiten bzw. Kampagnen-Themen zu erhalten. Erste Forschungsergebnisse zeigen, dass sogenannte Vornamen-Domains eine um 1,6 % geringere Löschwahrscheinlichkeit haben – die Herausforderung für »die Maschine« ist dabei, festzustellen, was tatsächlich ein Vorname ist (z. B. Peter und Petersilie) – wir verwenden dabei Listen der beliebtesten Vornamen lt. Statistik Austria.

Spannend ist auch der Blick auf die Website-Zertifikate: Hier zeigt sich nämlich, dass Domains, die etwa über ein SSL-Zertifikat verfügen, mit einer um 5 % höheren Wahrscheinlichkeit auch in einem Jahr noch aktiv sind.

KARL HEINZ WOLF
ist bereits seit 2007 Teil des R&D Teams bei **nic.at** und ist maßgeblich an der Produktentwicklung unserer gTLD-Software **tldbox** beteiligt.

RDAP: Challenge für alle gTLD Registries

Bis August 2019 muss für generische Top-Level-Domains (gTLDs) ein neues Protokoll in Betrieb gebracht werden. Das Registry Data Access Protocol (RDAP) soll das WHOIS-Protokoll ablösen, und wurde durch die ICANN für gTLDs verpflichtend vorgeschrieben. Dass das Whois-Protokoll nicht mehr den aktuellen Anforderungen entspricht, ist allerdings auch kein Wunder – immerhin wurde es bereits im Jahr 1982 erstellt! Das neue RDAP bringt zahlreiche Vorteile mit sich: So können etwa künftig auch nicht-ASCII-Zeichen (wie deutsche Umlaute) kodiert werden, die Ausgabedaten sind maschinenlesbar und der Zugriff auf die Domain-Daten findet über eine verschlüsselte Verbindung statt. Wir setzen RDAP für unsere gTLD **.versicherung** und alle unsere **tldbox**-Kunden wie etwa .berlin um.

MICHAEL BRAUNÖDER
gehört seit 2003 zum Team und arbeitet aktuell an mehreren Projekten, vorrangig im Bereich Data Warehouse.

Tagesaktueller »Domain-Wetterbericht«

Die in unserem Data Warehouse gesammelten Daten dienen als Basis für verschiedene Tools, die in unserem Unternehmen für Simulationen und Analysen herangezogen werden. So auch das Prognosetool, das aus der Vergangenheit lernt und eine tagesaktuelle Schätzung der bevorstehenden Neuregistrierungen und Löschungen liefert. Die Erfahrungen aus der Vergangenheit zeigen beispielsweise, dass Feiertage oder auch sehr gutes Wetter schlecht für unser »Domain-Geschäft« sind, da an solchen Tagen mit spürbar weniger Neuregistrierungen zu rechnen ist. Die Prognose ist somit auch eine wichtige Grundlage für unsere Budgetplanung. Wie treffsicher unsere Prognose-Tools inzwischen geworden sind, sehen Sie auf Seite 5.

Blockchain-basiertes Identity Mananagement

Spannend und zukunftsweisend zugleich: Die Datensouveränität (Self Sovereign Identity). Dabei entscheidet der Nutzer selbst, welche personenbezogenen Daten er wem anlassbezogen zur Verfügung stellt – die verschiedenen Identitäten können nicht miteinander in Verbindung gebracht werden. Eine Blockchain stellt dabei sicher, dass Identitäten nicht einfach zensuriert werden können. Die persönlichen Daten stehen in weiterer Folge ausschließlich den vom Nutzer zugelassenen Organisationen zur Verfügung, die manuelle Eingabe der Daten entfällt. Ein möglicher Anwendungsbereich in unserem Unternehmen ist die Ausstellung von Domainzertifikaten – relevant ist diese Technologie aber in jedem Bereich, in dem man sich anmeldet, ohne gleich seine kompletten Daten bekanntgeben zu wollen.

DIMITRIJ KLESEV
ist seit 2012 in der Backend-Entwicklung tätig und unterstützte das R&D Team bis Juni 2019, vor allem in seinem Fachgebiet »Blockchain«.

DOMAIN-VORSCHLAGSTOOLS
Unterstützung bei der Domain-Suche

Auf der Suche nach der passenden Domain steht man oft vor dem Problem, dass der gewünschte Wortlaut nicht mehr verfügbar ist. Sogenannte Domain-Vorschlagstools helfen in diesem Fall, Alternativen zu finden – allerdings sind hier abgesehen von Bindestrichen meist sehr wenig hilfreiche Vorschläge dabei. Grund genug für Philipp und Dimitrij aus unserem R&D Team, sich mit diesem Thema intensiv auseinanderzusetzen und unser eigenes Tool zu entwickeln.

Wie habt ihr das Projekt begonnen?

Dimitrij: Im ersten Schritt haben wir durch Data Processing bzw. Engineering die aktiven und gelöschten Domainnamen analysiert. Dabei wurden spezielle Algorithmen verwendet, um sie in einzelne Worte aufzuteilen oder Umlaute umzuwandeln. Wir haben uns die am häufigsten genutzten Wörter angesehen, die Längenverteilung bei Domains, aber auch die Zusammensetzung in Abhängigkeit bestimmter Faktoren. Ein Beispiel: Je länger eine Domain ist, desto eher enthält sie Bindestriche, was hinsichtlich Lesbarkeit eigentlich logisch ist, von uns aber zuerst mit Daten belegt werden musste.

Was sind aus eurer Sicht die größten Herausforderungen?

Dimitrij: Domains sind zu kurz, um morphologische Auswertungen zu erstellen und damit sämtliche Eigenschaften – etwa Syntax und Semantik – von Wörtern ermitteln zu können. Es wäre spannend, mit neuronalen Netzen zu arbeiten, was aber enorm komplex ist – so weit sind wir aktuell noch nicht. Eine Herausforderung ist auch, passende Algorithmen aus dem Natural Language Processing (NLP) Bereich zu finden.

Wie geht ihr damit um?

Philipp: Es ist grundsätzlich wichtig, Ziele zu definieren: Wann ist eine Lösung gut genug, wann hören wir auf? Eine zu 100 % perfekte Lösung die alles abdeckt werden wir nie haben – aber irgendwann muss man sich zufriedengeben und den nächsten Schritt machen.

Wie ist jetzt der aktuelle Stand?

Dimitrij: Wir können für einen Großteil der Anfragen Domains vorschlagen, die auch tatsächlich Sinn machen und »passen« – einerseits aus dem »Fundus« unserer sogenannten »Retro Domains« (gelöschte Domains), aber auch durch sinnvolles Einfügen von Bindestrichen und Umlauten. Das ist einmal der erste Schritt, wir testen noch viele weitere Ideen.

Wieso liefern herkömmliche Vorschlagstools oft so irrelevante Ergebnisse?

Philipp: Viele der Tools sind für die englische Sprache optimiert, das macht für uns natürlich nur begrenzt Sinn. Für Programme ist es auch schwierig, Sprachgebrauch und Kontext zu verstehen. Nutzergruppen sind heterogen, was für den einen sinnvoll ist, ist für den anderen total unbrauchbar. Es muss also ein Ansatz gefunden werden, der alle in einem gewissen Ausmaß zufrieden stellt.

Welche Rolle spielt in diesem Fall »machine learning«?

Philipp: Das ist für uns ein Werkzeug von vielen. Wir verwenden konkret ein Modell zum Trennen von Domainnamen in einzelne Worte, das wir auf Basis der bestehenden Bindestrich-Domains trainiert haben – das funktioniert bereits erstaunlich gut. Im Laufe der Zeit werden sicherlich noch weitere Modelle dazukommen.

Impressum: Medieninhaber & Verleger: nic.at GmbH, Jakob-Haringer-Straße 8/V, 5020 Salzburg. Konzeption und Redaktion: Lisa Hörbinger, Monika Pink-Rank; Fotos: Arman Rastegar, Gettyimages.com: Azret Ayubov. Verlags- und Herstellungsort: Salzburg, Juli 2019. Grafik: designkraft.at.

nic.at GmbH

Jakob-Haringer-Straße 8/V · 5020 Salzburg · Austria
T +43 662 46 69 -0 · F -29
pr@nic.at · **www.nic.at**

nic.at
.at Domain Registry by nic.at
.versicherung by nic.at
RcodeZero DNS by nic.at
TLD Box by nic.at