nic//report

07/2019

5th Edition // **RESEARCH & DEVELOPMENT**

nic.at

nic // report – Content

Dear reader,

Ever wondered where the many innovative research projects, ideas and numbers that surround our company originate? Or which tools inform our forecasts and strategic decisions? In this edition of the nic//report, we invite you to take a look behind the scenes at our R&D department and find out about the projects they are working on. Department head Alexander Mayrhofer also reveals how new ideas originate in the first place, how they become projects, how to keep yourself innovative, and what awaits us in the future.

The department continues to astonish even us with the things it creates and puts together – would you have thought that a dandelion seed head can illustrate the relationship between individual words used in domains? This report offers not only a wide-ranging insight into the R&D department, but a chance to expand your know-how, too – since those who can expertly reference terms like blockchain, RDAP and DNS over HTTPS, and are aware of current trends, will assuredly be rewarded with a knowing or appreciative glance at their next business appointment or dinner with friends. In this spirit, we wish you an enjoyable read!

Richard Wein und Robert Schischka
(Managing Directors of **nic.at**)

A LOOK BEHIND THE SCENES
Interview with Alexander Mayrhofer

To begin with, a general question: what are the responsibilities of the R&D department at **nic.at**?
Our activities can be divided into three areas: firstly, we work to develop and improve existing products, such as our gTLD registry, or **RcodeZero DNS**. The second area is focused on community and standardisation. At the international level, we play an active role in CENTR, and in the development of the internet by participating in the Internet Engineering Task Force (IETF). We organise events for the national community, like the DNSheads Vienna meetups. Our third focus area is research work – like, for example, gaining new insights from our data, as well as research on current topics such as identity management.

Our job is to work with highly complex things so that we can make use of the insights they offer and pass these on to our registrars and partners.

What projects are you working on at the moment?
Currently I'm working a lot on DNS Magnitude, a kind of statistical measure for the popularity of domains. Using a special algorithm, each domain is placed on a scale from 0 to 10 – providing a simple way to assess how intensively a domain is used. This can then be used by registrars, for example, for upselling – a popular domain should of course be appropriately safeguarded by products such as **Rcode ZeroDNS**. For instance, it would be very frustrating for both the customers and the operator if a popular online shop was temporarily unavailable.

Another major topic is encrypted DNS. Almost every action on the internet begins with DNS queries. This information is usually transmitted over networks in plain text. Two new protocols, DNS over TLS and DNS over HTTPS, aim to change this. However, DNS over HTTPS in particular also opens up the possibility that in the future the majority of DNS queries will only be responded to by a small number of large operators – a development that many are, at best, sceptical about.

Where do the ideas that are later turned into projects come from?
Many arise out of daily work, or simply when we have a little time to spend on new things. Talks with like-minded people who want to solve similar problems – so ideas from our community – are also often inspiring. But innovation does not always mean inventing something entirely new – most of the time, existing things are combined in new ways. An example of this is our data warehouse. We are just now starting to combine internal insights in new ways, and to offer them to registrars and the public.

How many of these ideas are actually implemented?
It's helpful here to use the concept of an innovation pyramid, based on a factor of 10: out of 3,000 ideas, 300 will be explored in more detail, of which 30 will become significant, specific projects, which can result in 3 completely new strategic directions for a business. So it's about a lot of trying things out, discarding ideas that don't work, and picking up something else to try.

What motivates you to stay curious? How does one stay innovative?
I like things that are initially very complicated, that I can simplify, so that they are useful and – this will sound a bit corny – just elegant. Our motivation comes from producing something useful that in the end can actually be applied and will contribute to the company's success. And I think that you stay innovative if you ask yourself, many times a day, questions like, "Can I make that better? Am I solving a problem? What is the benefit – how will users, registrars and our company profit?"

I'm proud that Google is currently implementing a new protocol for encrypted DNS that I standardised.

Is there not a danger of spreading yourself too thin?
Yes, definitely. You can't tackle too many things at once – the temptation is great, of course, because there are so many new ideas, and that's exciting. Innovation is always accompanied by the risk of "stranded work" – have courage and think outside the box, but be able to let projects go if they aren't heading in a useful direction.

Can you give us a glimpse of the future?
I can at least try. Machine learning will soon play an everyday role in ordinary businesses. We are already using it for forecasting. The challenge here will be how to make machines answerable, like people, for decisions they make – if an algorithm doesn't allow me to buy a train ticket, I will want to hear a justifiable reason.

Identity management will also be an important issue. The general public will be much more sensitive when it comes to data security and data sovereignty – a lot will happen in this area in the near future.

ALEXANDER MAYRHOFER
has managed the R&D department since 2002 and among other activities participates in the Internet Engineering Task Force (IETF). He often travels to speak at events, where both his talks and his eye-catching illustrations are always popular.

Alexander Mayrhofer's legendary "domain lagoon" illustrates the vastness of the ***.at*** *namespace – and that the number of domains actually registered can hardly get close to filling it. The "Isle of the Blessed" of 1.3 million registered* ***.at*** *domains is surrounded by the atoll of deleted domains, and the lagoon of WHOIS and EPP queries about free domains. This is encircled by the DNS sea of 300 million NX-domains a month – these are domain name queries received by our DNS servers for domains that do not exist. Out on the high seas lurk sea monsters in an ocean of unregistered domains which are never queried – making it clear that the potential for different* ***.at*** *domains is nowhere near exhausted.*

FACTS & FIGURES OF THE R&D DEPARTMENT

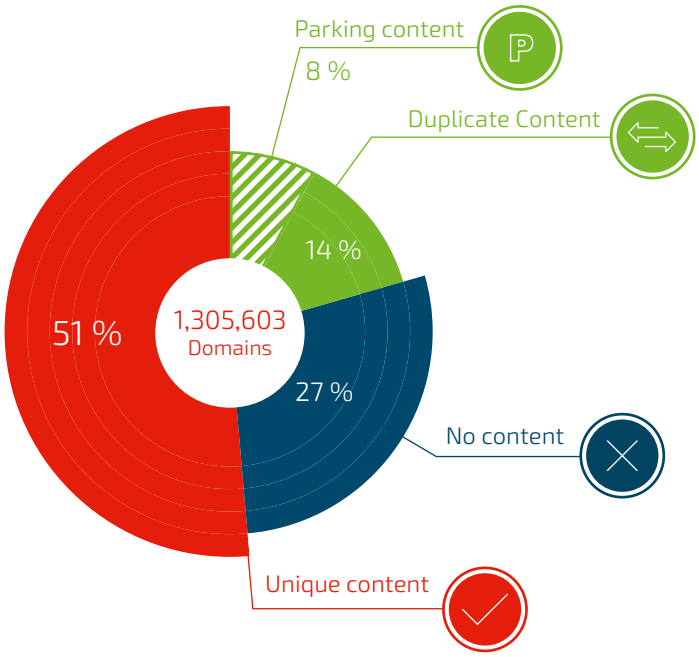
Top five first names in .at domains vs. most popular first names in Austria

	Domains	First names*
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

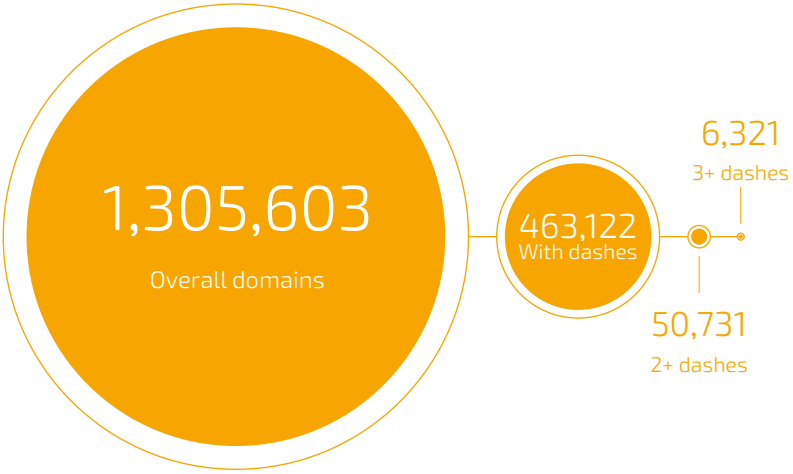
* Source: Statistics Austria

Websites behind .at domains



Domains containing dashes

(per June 25th, 2019)



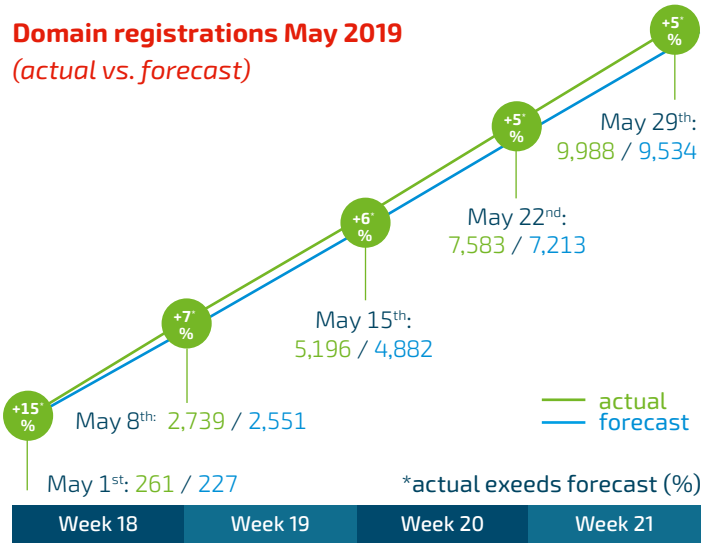
DNS magnitude per domain

*(The most popular websites in the **.at** zone)*

	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 registrations May 2019

(actual vs. forecast)



Top 30 words in .at domains

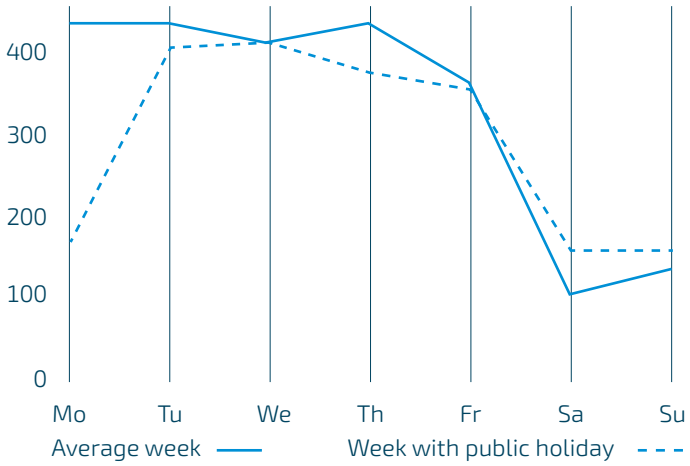
Number of possible .at domains*

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

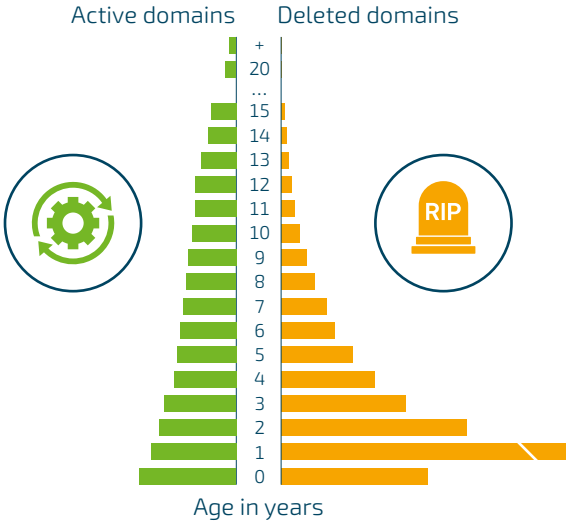
*at least

Domain registrations in the course of a week

(average, and when Monday is a public holiday)



Demography of the .at zone



Domains with a valid certificate

53.9 % of 1,306,503 domains *(as per June 12th 2019)*

Average shoe size

(of the R&D team)

CURRENT R&D PROJECTS:
Focal points for the employees

PHILIPP ADAM

has supported the team since 2019 as a Junior Research Engineer for the data warehouse. He is currently completing a master's degree at Vienna University of Economics and Business, specialising in data science.

Useful insights into domains

How are **.at** domains used? How quickly following domain registration does unique content appear on a site? Our **.at**-Webcrawler, which accesses the start page of the website behind every domain and analyses its content, can answer questions like these. For instance, whether a website with content is attached to a domain or not has an influence on the domain's longevity. This information is also interesting for registrars and partners, providing them with insights into their own portfolios, as well as potential upselling opportunities and campaign themes. Initial results from research show that domains including a first name are 1.6 % less likely to be deleted. Here, the challenge for the machine is to determine what really is a first name – e.g. telling the difference between "Peter" and "Petersilie" (the German word for parsley). To help it out, we use Statistic Austria's lists of the country's most popular first names.

A look at website certificates is also revealing: this shows that domains with an SSL certificate are 5 % more likely to still be active after a year.

KARL HEINZ WOLF

has been a member of the R&D team at **nic.at** since 2007, and is heavily involved in product development for our gTLD software **tldbox**.

RDAP: A challenge for all gTLD registries

A new protocol must be implemented for generic top-level domains (gTLDs) by August 2019. The Registry Data Access Protocol (RDAP) will supersede the WHOIS protocol, and has been made obligatory for gTLDs by ICANN. The WHOIS protocol was set up in 1982, so it comes as no surprise that it no longer meets today's requirements. RDAP offers a large range of benefits. For example, in future it will be possible to write code using non-ASCII-characters (such as German vowels with umlauts), output data will be machine-readable, and domain data will be accessed via an encrypted connection. We are implementing RDAP for our **.versicherung** gTLD and all of our **tldbox** customers, like **.berlin** for example.

MICHAEL BRAUNÖDER

has been a member of the team since 2003, and is currently working on a number of projects, primarily for the data warehouse.

Daily weather report for domains

The data collected in our data warehouse form the basis for various tools that are applied in-house in simulations and analyses. This includes the forecasting tool, which learns from the past and provides a daily prediction of new registrations and deletions. For example, experience from the past shows us that public holidays and very good weather are bad for our domain business – on these days significantly fewer new registrations can be expected. The forecasts therefore also provide important input for budget planning. See page 5 for a breakdown of just how accurate our forecasting tools have become.

Blockchain-based identity management

Self-sovereign identity is both exciting and forward-looking. It allows the users to decide for themselves which personal data they make available, depending on the situation, by ensuring that the various identities cannot be connected with each other. A blockchain guarantees that identities cannot simply be censored. The personal data are then available only to the organisation granted access by the user. Manual data entry is no longer necessary. One potential area of application in our business is issuing domain certificates – but the technology is relevant wherever users want to register online without sharing all of their data in doing so.

DIMITRIJ KLESEV

has worked in backend development since 2012, and supported the R&D team until June 2019, principally in his specialist area of blockchain technology.

DOMAIN SUGGESTIONS TOOLS
Domain search support

When looking for the right domain, the problem can often be that the desired domain name is no longer available. Domain suggestion tools help users to find alternatives – but apart from inserting dashes, they usually produce few genuinely helpful ideas. This was reason enough for Philipp and Dimitrij from our R&D team to tackle the problem head-on, and develop a proprietary tool.

How did you get started?

Dimitrij: The first step was to analyse active and deleted domain names using data processing and data engineering. We used special algorithms to break them up into individual words and to convert them to include umlauts. We looked at the most frequently used words, the distribution of domain name lengths, and the composition of names in relation to certain factors. An example: the longer a domain name, the more likely it is to contain dashes – this is logical with regard to readability, but we had to demonstrate it with data.

From your perspective, what are the biggest challenges?

Dimitrij: Domain names are too short to be able to analyse them morphologically and derive all of the properties of each word, such as syntax and semantics. It would be exciting to work with neural networks, but also extremely complex – we haven't quite got that far yet. Another challenge is finding suitable algorithms for natural language processing (NLP).

How are you approaching the problem?

Philipp: First of all, it's important to define the objectives: when is a solution good enough, where do we stop? We will never have a 100 % perfect solution that covers everything – but at some point you have to be satisfied and move on to the next step.

What is the current status?

Dimitrij: We are able to suggest domains that actually make sense and are a good fit for a majority of enquiries – either from our store cupboard of "retro domains" (deleted domains), or by inserting dashes and umlauts. This is just the first stage – we are continuing to test many other ideas.

Why do conventional suggestion tools often provide such irrelevant results?

Philipp: Many of the tools are optimised for English – obviously that doesn't help us much. It's also difficult for programs to understand language use and context. User groups are heterogeneous – what makes sense for one can be completely useless for others. This means we need to find an approach that satisfies everyone to a certain degree.

What role can machine learning play?

Philipp: It's an extremely useful tool for us. We currently use a model, that we have trained on the basis of existing domains containing dashes, to split up domain names into individual words – it already works astonishingly well. Over time further models will definitely be added.

Imprint: Owner and publisher: nic.at GmbH, Jakob-Haringer-Straße 8/V, 5020 Salzburg, Concept and editorial: Lisa Hörbinger, Monika Pink-Rank; Photos: Arman Rastegar, Gettyimages.com: Azret Ayubov. Place of publication: Salzburg, July 2019. Graphics: 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**