

To: GNSO Council

My name is Vassil Petev, a proud Bulgarian citizen, a Unit Manager in an established, leading, and world-known Bulgarian IT company, Telerik Inc. ([www.telerik.com](http://www.telerik.com))

I am writing to you on behalf of above parties to express my opposition to a rule in the current version of Module 2 of the Draft Applicant Guidebook, and specifically, the section on "Outcomes of the String Similarity Review".

I would also like to communicate my support for the GNSO Council request that the Applicant Guidebook section on "Outcomes of the String Similarity Review" be amended to allow applicants to request an Extended Review under applicable terms similar to those provided for other issues such as "DNS Stability: String Review Procedure". I also agree that a section be added on "String Similarity – Extended Review" that parallels other such sections in Module 2.

I hereby present to you my view and understanding of the application process for cases where strings are "confusingly similar" to other TLDs and/or ccTLDs, and request that these views are included in your recommendations to Kurt Pritz, the members of the ICANN New GTLD Implementation Team, and the ICANN Board. My views and understandings are backed up with two real-life examples, the first one concerning my country, and the second one covering a case with the company I work for.

## Contents

The case of the Cyrillic ccTLD for Bulgaria (*.бг) and the Latin IDN for Brazil (*.br) .....	2
The population that speaks Cyrillic languages.....	2
The population that speaks Latin languages.....	3
Brazil's IDN vs. Bulgaria's IDN .....	3
The Extreme Case .....	3
ICANN Staff's reasoning on *.бг.....	4
More reasons on why *.бг is Bulgaria's IDN .....	4
The case of the Cyrillic gTLD *.ком (for компания) and the Latin gTLD *.com (for company) .....	5
Putting *.ком in context .....	5
*.com – the most widely used gTLD in the world.....	5
What does *.ком stand for, and why it is so special to ICANN staff? .....	5
Fraud and companies operating in Cyrillic speaking countries .....	6
More reasons for the Cyrillic *.ком gTLD .....	6
Why this matters to me: The history of the Cyrillic alphabet.....	7
Similarities and differences between Cyrillic and Latin characters .....	8

## The case of the Cyrillic ccTLD for Bulgaria (\*.бг) and the Latin IDN for Brazil (\*.br)

Bulgaria is a nation which is directly impacted by the current automatic disqualification when strings are "confusingly similar" to other TLDs, in this case a ccTLD, and already has been declined twice (in late 2009, and in May of 2010) to register the \*.бг Cyrillic IDN on the premise that it looks confusingly similar to Brazil's \*.br ASCII TLD.

### The population that speaks Cyrillic languages

Although a Latin-speaking user can certainly find these strings quite similar, a Cyrillic speaking person will know which one is which. The Cyrillic letter **б** does not look like a **b** (see appendix for comparison of the Latin and Cyrillic alphabets); it actually looks much more like the number **6**, however every person who speaks a Cyrillic language will recognize the difference between the two letters and the number, especially when put into context (more on this later). This case is even more obvious in hand-written fonts: *.δz vs .br*

The difference between the subsequent top-level domain letters **г** and **r** are not as noticeable in regular fonts, but are very noticeable in hand-written and italic fonts (see the heading in this section as an example). Still, a person who knows a Cyrillic language will know the difference.

So, the problem here is with the population that speaks Latin languages who find these strings confusingly similar.

## The population that speaks Latin languages

A major point that ICANN is missing in their current evaluation criteria for confusingly similar strings is that they do not review the TLDs, especially IDNs, in the **context they will be used in**. When reviewing an IDN in context, the evaluation of the string (and its alphabetical differentiation) becomes much easier. As an example, let's look at how my company's domain would look like in Latin and Cyrillic IDNs:

- telerik.br
- телерик.бр

I doubt that someone will mistakenly take one for the other. Still, let's analyze this in more detail and review some extreme similarity cases.

## Brazil's IDN vs. Bulgaria's IDN

The main reasons that differentiate Brazil's IDN from the Bulgarian IDN are:

1. A URL consists of a top-level domain *and* a second-level domain. Since **.бр** and **.br** are just the top-level domains, they are meaningless without a second-level domain. When comparing full URLs, the difference between the two is exceptionally obvious.

*Example:* **telerik.br** and **телерик.бр**

2. Brazil uses three tier domains (host+ gTLD+ ccTLD), whereas Bulgaria uses two-tier domains (host+ ccTLD), which makes the visual gap between the two even larger. As a result, a Brazilian user looking at a Bulgarian URL will know right away that this is not a Brazilian domain, even if the host uses the same letters.

*Example1:* **Vivo** is one of Brazil's mobile network operators. Their site is **vivo.com.br** which in Bulgarian would be **виво.бр**. There is no resemblance between the two.

*Example2:* An imaginary company is called American Electric, and its main domain is **ae.com**. Its Bulgarian domain would be **ae.бр**, which does not resemble **ae.com.br**, even though the host is exactly the same. Even if Bulgaria starts using three-tier domain names (host+gTLD+ ccTLD), this URL will look like **ae.ком.бр** (*company* (ENG) = *компания* (BG)), which is also decidedly not the same as the Brazilian domain.

## The Extreme Case

IMPORTANT NOTE: The analysis below is excessive, and this is on purpose, because it could happen. It raises the importance of having regulations in the case that such situations arise in the future. This analysis presumes that Brazil uses two-tier domain names (host+ccTLD), and that there is a company with a domain string that is exactly the same in Cyrillic and Latin languages.

1. If a non-native English speaker (such as a Frenchman or a Spaniard) sees **ae.бр**, but knows the context where the URL is used/mentioned, s/he will most probably know that this is a Cyrillic/Bulgarian domain. No action here.

2. If a non-native English speaker (such as a Frenchman or a Spaniard) or a native English speaker sees **ae.6r** without knowing the context where the domain is used/mentioned, s/he *may* think that it is in Latin.

*In such cases, regulation should be in place to control the use of these strings and to ensure that a single registrant owns visually similar domains.*

*In addition, browser vendors need to update their error message in case **ae.6r** is entered in Latin letters in the browser, and there is no such domain. The error message should reflect that the domain may be in Cyrillic. Here is an example for a possible error message:*

#### **Server not found**

**Firefox can't find the server at www.ae.bg.**

- Check the address for typing errors such as **ww.example.com** instead of **www.example.com**
- Check the address for being in Cyrillic such as **ae.6r** instead of **ae.br**
- If you are unable to load any pages, check your computer's network connection.
- If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web.

### **ICANN Staff's reasoning on \*.6r**

ICANN staff's reasoning for declining Bulgaria is that "internet is a world resource and uniqueness is most important." However, its decision will have an impact on at least 7 million Bulgarians, not to mention their relatives and the Bulgarian-speaking population around the world. In addition, with the IDN ccTLD Fast Track Process ICANN wants to open the Internet to languages based on scripts other than Latin in order to make it more accessible, but at the same time impose limitations on its openness, thus effectively contradicting itself.

### **More reasons on why \*.6r is Bulgaria's IDN**

There is no doubt in my mind that the \*.6r IDN is the most appropriate IDN for Bulgaria, for the following reasons:

- Last year the Bulgarian government ran a survey on the IDN that Bulgarians would like to see in use. The Bulgarians chose \*.6r and this IDN won hands down;
- Bulgarian nationals will not have to learn a new abbreviation. The **BG** abbreviation has been used for years to represent **Bulgaria**; one can find it in the Bulgarian currency, passports, car licenses, stickers, maps, etc. The \*.6r IDN is the same abbreviation, but in Cyrillic script;
- The \*.6r IDN feels and writes natural to Bulgarian-speaking population;
- Typing Cyrillic requires Bulgarian language and keyboard to be installed on the machine, so one cannot easily type Latin while in Cyrillic mode. Typing Portuguese requires either a Brazilian or a US language and keyboard.

To close this case, I would like to add that comparing Cyrillic script to Latin script is like comparing apples and oranges. As the above reasoning shows, such comparison can be disputed; with small set of regulation rules in place, everyone will be happy.

## The case of the Cyrillic gTLD \*.ком (for компания) and the Latin gTLD \*.com (for company)

In this case, it is not only Bulgaria, but more than 10 Cyrillic speaking countries (incl. Russia) that are directly impacted by the current automatic disqualification process when strings are "confusingly similar" to other gTLDs. Here, ICANN plans to ban the Cyrillic gTLD \*.ком (for "компания" in Russian and Bulgarian or "company" in English), because it looks confusingly similar to the current \*.com gTLD.

The impact of this decision, if made final, will be on **millions of registered companies** in these countries, and on at least **tens of millions of people** who are used to see and use \*.com, and who believe that the natural Cyrillic twin of \*.com is \*.ком.

### Putting \*.ком in context

Once again, ICANN's current evaluation criteria for confusingly similar strings do not cover the case of how the gTLDs will be used, i.e. in what **context**. When reviewing a gTLD in context, the evaluation of the string (and its alphabetical differentiation) is much clearer. As an example, let's look at how my company's domain would look like in the new Latin and Cyrillic gTLD:

- telerik.com
- телерик.ком

I doubt that someone will mistakenly take one for the other.

### \*.com – the most widely used gTLD in the world

**.com** is the most widely used gTLD in the world, and for this reason alone it makes sense for ICANN to find an excellent Cyrillic counterpart for the \*.com TLD, which to be as close as possible in terms of pronunciation (to avoid unnecessary confusion), AND have the same (or very similar) meaning. The \*.ком Cyrillic gTLD not only sounds the same in Cyrillic, but its meaning is the same as its Latin equivalent! I really do not understand why ICANN would like to push yet another gTLD to the millions of companies registered in Cyrillic speaking countries, and their citizens, when they have the perfect match.

### What does \*.ком stand for, and why it is so special to ICANN staff?

I am not sure, but ICANN staff seems to have its reasons. Searching <http://wordfamous.com/> for words that start with "kom" (the exact search string I used is "kom\*" without the quotes) revealed 12 matches in the English language, which are actually 7 words:

1. [komatik](#) and [komatiks](#)
2. [kombu](#) and [kombus](#)
3. [komissar](#) and [komissars](#)
4. [komitaji](#) and [komitajis](#)
5. [komondor](#) and [komondors](#)
6. [komondorock](#)
7. [komondorok](#)

I then went to <http://www.wordreference.com/> and checked for words that start with “kom” in French, Italian, and even Portuguese languages, but found just a few more results.

I need to disclose that I am not an expert in these languages (although I speak ), so my quick research should be taken just as a pointer, rather than the truth. But, if indeed there are just a handful words starting with this Latin string, why is \*.kom so special, and why ICANN wants to protect it? It seems ICANN is ready to protect a few rarely used words, but sacrifice the rights of a few hundred million people and millions of companies in Cyrillic speaking countries over them.

May be ICANN is worried that someone may apply for the \*.kom registration, which (if approved) to lead to confusion. This situation is not possible, because ICANN is the only organization that can actually approve this application; simple screening of the application (where ICANN is very proficient at), should be enough to decline the application, and avoid such disorder.

### Fraud and companies operating in Cyrillic speaking countries

To avoid fraud and cybersquatting, ICANN should provide the means to secure \*.com domain owners and their domains in the Cyrillic version of .com, i.e. provide them with the right of first say when their trademark is involved. Ruling against the Cyrillic \*.ком gTLD will not stop this process. For example, the company I work for, Telerik, should not be forced to buy Cyrillic domains just to prevent fraud by others trading on our good name.

Telerik wants the have the option to activate its own \*.ком Cyrillic equivalent, because it is a privilege, it is prestigious, and people will be able to find us in their own language. Telerik is an excellent place to work for, and it has won two significant awards<sup>1</sup> proving that. Many Bulgarians crave for working in Telerik, and they want to take part of the company that is making a change in the IT industry. They will be obliged to work for a company that also owns its own Cyrillic **телерик.ком**. We want the ability to activate our own \*.ком domain and will pay for that privilege.

### More reasons for the Cyrillic \*.ком gTLD

There is no doubt in my mind that the \*.ком gTLD is the most appropriate \*.com equivalent, for the following reasons:

- \*.ком is the spot-on string in Cyrillic for companies. It stands for “company”, it sounds like \*.com, and it is in Cyrillic. ICANN staff should not automatically disqualify the .ком Cyrillic application on the sole premise of similarity;
- \*.ком may look like \*.com, but it is very different. This case is even more obvious in hand-written fonts: *.ком* vs *.com*
- Although a Latin-speaking user may certainly find these strings similar, a Cyrillic speaking person will know which one is which.

---

<sup>1</sup> 2007/2008 Best Employer in Central Eastern Europe, 3<sup>rd</sup> place  
2008/2009 Best Employer in Bulgaria 2009, 2<sup>nd</sup> place  
<http://www.telerik.com/company/careers.aspx>

- I have always been a proponent of free speech, free markets, etc. It is my belief that in order to find the best solution, one should propose a few alternatives which to provide to people to play with. People are intelligent and they will tell you what they like. Imposing restrictions before actually surveying what they need deeply intervenes with their right of choice.

## **Why this matters to me: The history of the Cyrillic alphabet**

To finish off, I would like to give you a little background on the Cyrillic alphabet.

The Cyrillic script is an alphabet developed in the 9th century by two brothers, Cyril and Methodius, who were later on venerated in the Eastern Orthodox Church as saints. The Cyrillic alphabet was first adopted by Bulgaria, my home country, and because of that Cyrillic is believed to be a Bulgarian alphabet, although this is debatable. The Cyrillic script is used in the Slavic nations of Belarus, Bosnia, Bulgaria, Russia, Serbia, Macedonia, Montenegro, and Ukraine, and in the non-Slavic nations of Moldova, Kazakhstan, Uzbekistan, Kyrgyzstan, Tajikistan, Tuva, and Mongolia. With the accession of Bulgaria to the European Union on 1 January 2007, Cyrillic became the third official alphabet of the European Union, following the Latin and Greek alphabets. It is also one of the few alphabets that has its own holiday (May 24th), which is celebrated internationally.

In conclusion, I encourage the Council to send its request to ICANN staff, with inclusion of the examples I have described here. Further, I strongly encourage ICANN staff and Board to follow the advice of GNSO Council to create an extended evaluation for string similarity.

I will be glad to assist you if my services and knowledge are required to protect the rights of the people and companies that do business in Cyrillic-speaking countries.

I thank you for your prompt attention to this request.

Yours truly,

Vassil Petev  
Unit Manager, Telerik Inc.  
Bulgaria

## Similarities and differences between Cyrillic and Latin characters

Legend:

Green background – exact or very close matches, regardless of the font used
Grey background – may look similar, depending on the font used. Segoe Script used to illustrate the differences.

Capital Letters (upper case)			Lower Case		
Cyrillic	Latin Equivalent	Latin Look-a-likes	Cyrillic	Latin Equivalent	Latin Look-a-likes
A	A		a	a	
Б	B		б	b	6 (number six)
В	V	В	в	v	b
Г	G		г	g	r
Д	D		д	d	
Е	E		е	e	
Ж			ж	-	
З	Z	З (number three)	з	z	3 (number three)
И	I	И	и	i	n
Й	Y		й	y	
К	K		к	k	
Л	L		л	l	
М	M		м	m	
Н	N	Н	н	n	H
О	O		о	o	
П	P		п	p	
Р	R	Р	р	r	ρ
С	S	С	с	s	c
Т	T		т	t	m
У	U	У	у	u	y
Ф	F		ф	f	
Х	H	Х	х	h	x
Ц			ц		
Ч		Ч (number four)	ч		4 (number four)
Ш		Ш	ш		
Щ		Щ	щ		
Ъ			ъ		
Ь			ь		
Ю			ю		
Я		Я	я		