TRADEMARK INFRINGEMENTS IN THE DOMAIN ".CZ"

Tomáš Gongol

Introduction

This article deals with finding and evaluating the extent of trademark infringements in the field of domain ".cz" (further referred to as Czech domain). Not only in the Czech legal environment, the question of disputes between intellectual property rights (esp. trademarks) and domain names has traditionally been included in the interpretation of information technology law (Polčák et al., 2018; Lloyd, 2011) or internet law (Jansa et al., 2016; Edwards & Waelde, 2009). Trademarks have a number of functions in the market economy that are described in a number of professional publications (e.g. Horáček et al., 2017) and also extended by follow-up judicial practice. The trademark is an important business identifier for entrepreneurs. It reinforces sales of goods and services on the market, therefore, the entrepreneurs invest considerable financial resources into promoting it (see, for example, Crass et al., 2019). Its basic function is to distinguish the products or services of one trader from the products and services of another and it protects consumers from misleading (Lukose, 2013). It is also an effective tool for providing information on the market (Griffiths, 2008; Burmann, 2017) and helps its owners to obtain and maintain a position in the competition (Slováková, 2006; Munková et. al., 2012). The trademark infringement that is left without an adequate response can lead to the reduction of its distinctive character, and ultimately to its demise. This is true not only in the material world, but also in the digital environment (see Merges et al., 2012). The use of the trademark as a domain name can have a negative impact not only on advertising, but also on the business strategy of its owner, so it is important to address the question of the actual state of its abuse. For such examination, it is necessary to use an interdisciplinary approach involving not only the area of the law but also of marketing, the economics of information, information technology, cybernetics and statistics. Thousands of disputes being resolved worldwide whether judicially or the alternate way suggest the extent of the conflict of trademarks vs. domain names. There has been an extensive publishing activity to this issue. It addresses both, the nature of dispute procedure, e.g., Gongol (2014), Pelikánová (2012), Jansa et al. (2016), Werra (2016), and analyses of specific disputes resolved this way, e.g., Bettinger (2005), Merges et al. (2012), Gongol (2013a). There also known more ways of trademark infringement on the Internet, part of which are the use of trademark as a keyword in the search engines (referred to by Gongol, 2013b; Janis & Dinwoodie, 2007; Gielen, 2010; Senftleben, 2012; Oullette, 2014), in metadata of websites, e.g. auction portals (referred to by Otim & Grover, 2010; Saunders & Berger-Walliser, 2011; Gongol, 2016). In this context, however, it is possible to ask several research questions that extend the explored region further and deepen the real condition in the Czech domain:

1. Is the phenomenon of the trademark infringement a rather minor issue, which concerns only a fraction of domain names and websites, or is it a widespread practice on the Internet?
2. Do domain names or websites that violate trademark rights share similar characters?
3. Is it possible to automate the process of searching for trademark infringements or is it possible to find general rules that can be a valuable help in this process?

To date, there is not known any relevant study that would have dealt with these issues more deeply and provided even an approximate quantification or methodology. Some of the sub-aspects are dealt with by the older study by Branthover (2002) that is focused on the

 DOI: 10.15240/tul/001/2019-4-011

results of the disputed proceedings, or the success of the complainant in the correlation with the selection of the court, which is further followed by a newer study focused on the forum selling by Klerman (2016). From newer works can be mentioned Visserse et al. (2015), which focuses on the analysis and detection of a narrower circuit called Parked domains. Similarly, a study focused on malware and phishing domains from Korczynski et al. (2017), which follows the previous analysis by Halvorson et. al. (2015). Analytic and statistical reports are also available at administrators of both national and generic domains. The basic statistic overview is also provided by the administrator of the Czech domain CZ.NIC. on its website. In order to be able to answer the above questions in general terms at least, it is necessary to link the results of the analysis of the decision-making practice given in the previous work of the author (Gongol, 2012) to the real situation on the Internet. It is necessary to examine the websites, their content and domain names and put them into correlation with existing trademarks. In this article the attention will be focused on the Czech domain, specifically on measuring the amount of trademark infringements in automotive industry in which a considerable amount of investments is carried out. The automotive sector is widely represented on the Internet. Therefore, the websites of automakers are often searched among the Czech population which in fact creates good conditions for research. In the automotive sector will be examined a series of characters that occur in cases of trademark infringements or, on the contrary, lead to confirmation that the website does not abuse a trademark.

1. Methodology

On the basis of the decision-making practice of the World Intellectual Property Organization (WIPO) and the Czech Arbitration Court attached to the Economic Chamber of the Czech Republic and the Agricultural Chamber of the Czech Republic (CAC), can be identified different characters that lead to the confirmation or refutation of the conclusion on trademark infringement on a specific website on particular domain name. For selected characters, it is necessary to define clear semantics and rules so a computer algorithm will be able to decide whether a specific domain name or a website is in compliance with them or not. Then, within the set of monitored domain names and trademarks, it will be possible to statistically determine which characters are typical for both types of domain names, those that do abuse trademarks and those that do not.

First of all it is necessary to define the input sources. As mentioned, the attention will be focused on the area of the Czech domain, therefore, it is necessary to get a current list of domain names in the .cz domain, which is about to be analyzed. As a source of trademarks will be used the database of the Industrial Property Office (IPO). Given the need of finding out which domain names are relevant to the particular trademark, it opens the possibility of detecting the extent in which the words contained in the trademark appear in the text of the domain name. This way it is possible to link one with the other. The mere determination of whether a trademark is part of the text of a domain name is not entirely accurate. For example, the trademark “audi” is part of the text of the domain “audio.cz” although there is no relationship between the two. In the article, therefore, will be used a more precise mechanism using Czech dictionary and Thesaurus of Czech words. The textual representation of domain names and trademarks will be separated using a thesaurus on individual words and the coincidence between trademarks and domain names will be further examined at the level of words. The above example of *audi* vs. *audio.cz* will have already traversed correctly. Upon found relationships between domain names and trademarks will be performed fundamental analysis, which will provide answers to questions such as, how many percent of domain names in the .cz domain is related to a trademark, how many domain names on average are related to one trademark or which trademarks are the most represented on the Czech Internet. It is possible to search the words constituting the trademark not only in the text of the domain name but also in the text of a website. Such comprehensive research, however, is outside the scope of this article. We will work with the assumption that the trademark is represented in some way in the text of the domain name, as this practice is more or less taken into account by Internet search engines (Janouch, 2014); Internet users, while searching for pages that relate to the particular trademark, may type the text of the trademark directly as a domain name

(example: cocacola.cz, coca-cola.cz, skoda-auto.cz, etc.).

The sector of the automotive industry will be studied further in detail. In order to ascertain the effect of the occurrence of particular characters on the determination whether a trademark is infringed on a website or not, the method of logistic regression (Kleinbaum et al., 2010) will be used. In order to perform analysis, it is necessary to go manually through the monitored domain names (approximately 1600 domain names) while taking into account the decision-making practice (WIPO and CAC) and accordingly divide them into two categories: those in the conflict of law (trademark infringing) and non-conflictive (no infringement) – hereinafter referred to as collision/non-collision. Using defined explanatory characters, the category of collision/non-collision is going to be explained via regression analysis.

2. Data Acquisition and its Processing

2.1 List of Domain Names

For the acquisition of the current list of domain names .cz is the best to get the information directly from the database administrator of the domain CZ.NIC, Interest Association of Legal Entities. The administrator, however, refuses to provide the data, albeit only for research purposes. For this reason, a list of domain names, current to January 2015 and provided by CESNET, Interest Association of Legal Entities serves as a base data file. Although, the disadvantage of the list is its incompleteness. There are 517,655 domain names out of 1,178,891, which makes approx. 44% of the total number of registered domain names at CZ.NIC in that time. To refine the list of current domain names there were used additional sources of domain names listed on *czdomeny.cz* and *domainpunch.com*. This resulted in an overall list containing 568,272 Czech domain names, which are used further on.

2.2 List of Trademarks

In another part of the article, there is for the selected trademarks detected a link between relevant domain names and content of websites connected to these domain names, which will then be used for the analysis of so-called Characters. On the territory of the Czech Republic, trademarks listed in Section 2 of Act No. 441/2003 Coll. on trademarks, shall enjoy the protection of trademark. The current list of national trademarks is publicly available on the IPO website. From the database of IPO were downloaded 474,222 unique lexical trademarks by using a computer program. Trademarks are made up of phrases where each word is separated by a space (e.g. *audi spare parts* or *genuine škoda accessories*). Registered trademarks consist of 1.72 words in average. The following Tab. 1 shows the layout of a number of trademarks for a specific number of words.

Because of the computing requirements (search for links between approximately 470 thousand trademarks in relation to 560 thousand domain names) for the purpose of this article,

Tab. 1: The number of trademarks for a given word count

Word count	Number of trademarks	%	Sum %
1	266,697	56.24%	56.24%
2	133,919	28.24%	84.48%
3	44,982	9.49%	93.96%
4	16,378	3.45%	97.42%
5	6,438	1.36%	98.78%
6	3,032	0.64%	99.41%
7	1,295	0.27%	99.69%
8–1,000	1,481	0.31%	100.00%
Total	474,222	100.00%	

Source: own

Tab. 2: Words obtained from dictionaries and trademarks

The source	Word count	%
The word obtained exclusively from the trademarks	286,541	49.83%
The words obtained from dictionaries and lists of municipalities and names	253,311	44.05%
The words contained in both sources	35,179	6.12%
Total	575,031	100.00%

Source: own

there are only used such trademarks that consist of 5 and less words – these make up 97.42% of all trademarks. For the same reasons this work focuses only on those trademarks having 4 – 160 characters (these make up over 98% of all trademarks). Trademarks may contain words of Czech language, names of people, etc. For such trademarks and particularly domain names that are paired to them, it is difficult to determine whether the domain name refers to the meaning of the word commonly used in the Czech language or to trademark rights (e.g. trademark registered in IPO "Osamělý vlk"). In this article, therefore, we will focus only on trademarks which have more distinctive eligibility and do not contain only Czech words of generic or descriptive character (especially because of the power of efficiency of trademark on the Internet). From a total of 474,222 trademarks, 361,007 of them meet these criteria (76.13%).

2.3 Other Resources

In order to determine whether a trademark contains Czech words, it is necessary to use the current Czech Dictionary or the Czech Corpus. For these purposes, it is used the Czech language corpus of SYN2010 from the Institute of the Czech National Corpus. When analyzing the relationship of domain names and trademarks it is necessary to split the domain name into individual words, in order to determine compliance with the trademark (e.g., for "mojeaudi.cz" the domain name needs to be split into words "moje" and "audi" to determine the continuity of the audi's trademark). Domain names, however, not only consist of Czech words but also of English words. They commonly contain the names of municipalities or first and last names of people (e.g., myaudi.cz – a composition of the English word "my" and the trademark of "audi"; bartekskoda.cz – a composition of last name "Bártek" and the trademark of "skoda"; audipraha.cz – a composition of the trademark "audi" and city name "Prague", etc.). The total work corpus that is used to split domain names to single words, contains the Czech language corpus of SYN2010 (Institute of the Czech National Corpus), the English language corpus (GNU-FDL English-Czech Dictionary), list of Czech municipalities, list of Czech names and surnames, words found in trademarks of IPO database. Because of the computing requirements, words of less than 4 characters are taken off, unless those are prepositions or conjunctions of Czech or English language. The total resulting corpus contains 575,031 words, of which about half comes from neologisms extracted from trademarks.

3. Analysis of Correlations between Domain Names and Trademarks

With the available resources can be done a general analysis of correlations between trademarks and domain names. It is necessary to answer the question of which domains are relevant for a given trademark. As an example the domain "servisrenault-stredoceskykraj.cz" is relevant to the Renault trademark; the domain "vrakoviste-fiat-renault.cz" is relevant to both Renault and Fiat trademarks; however, domain "audiostezka.cz" is only relevant to the trademark "Audiostezka" but not to "Audi" trademark. As you can see in the examples, the correlation between trademarks and domain names is generally N : N, hence one trademark may be relevant in multiple domain names, and a single domain can be relevant to more trademarks (e.g., the above referred "vrakoviste-fiat-renault.cz"). First, there will be used a simple way of finding a domain name

for a trademark by a direct search for parts of the trademark in the text of the domain name. The algorithm first splits the trademark into individual parts according to a space (and other separators such as comma, semicolon, etc.) and then it searches whether the trademark is part of the text of a domain name. For example, a trademark "Potrefená husa" is first split into parts "potrefena" and "husa". Then the algorithm goes through all of the domain names and looks for such text, which contain both of the words. Relevant domains such as "potrefenahusa-design.cz", "potrefenahusazlin.cz" will so end up in the final result. However, this algorithm does not always work correctly. For example, the algorithm correctly finds domain names containing the trademark "audi", such as *audi.cz* or *chiptuning-audi.cz*, but it also finds domain names that have nothing in common with Audi trademark, such as *aaudio.cz* or *absoluteaudio.cz* and others. Another, more complicated algorithm can tackle this problem using the Czech and English corpus to prevent such misleading match. The following Tab. 3 shows the result of the above-described algorithm process.

Tab. 3: Parts of trademarks in the text of domain names

	Number of TM	%
The TM is included in the text of a domain name	61,482	17.03%
The TM is NOT included in the text of a domain name	299,525	82.97%
Total	**361,007**	**100.00%**

Source: own

From the above, it is evident that for 61,482 (17.03%) trademarks, there exists a domain name that consists of words contained in the trademark. Because we focused only on trademarks with more distinctive eligibility, which could potentially lead to a confusion of their meaning, we can assume that the domain names that were found can actually be relevant to those trademarks. As mentioned above, the simple text match algorithm has its weak spots and is shown here only for comparison. In the following text, there will be used a more complex algorithm, which is more accurate in the process of determining domain names for given trademark. The algorithm will use the Czech and English corpus, which shall split the trademarks and domain names into words and the match will be determined at the level of words. For the above mentioned "absoluteaudio.cz" the algorithm divides the domain name into English words "absolute" and "audio", making them no longer bound to the trademark Audi. The basic definition of the used algorithm:

1. For each domain name is searched a sequence of words based on the corpus, which a domain name consists of (for example, the words "moje" and "audi" put together create a domain name "mojeaudi.cz"). There is a function within the algorithm, which finds phrases that form a given domain name. The algorithm maximizes the function to find the best combination of words.
2. Trademarks are split into words. Characters such as space, comma, dot, etc. are used as separators.
3. By splitting domain names into words there are found relevant domain names for all 361,007 trademarks. It is essential that words contained in a trademark must also be part of the above-described domain name division into words.

The results of the above algorithm are shown in Tab. 4.

By using a more accurate algorithm was found that 43,436 (12.03%) trademarks are linked to one or more of the Czech domain names. With knowledge of the links between trademarks and domain names, it is possible to specify how many Czech domain names were made with the intention to benefit from the existence of some of the trademarks. Of a total of 568,272 Czech domain names, 123,050 (21.65%) domains are bound to a trademark, see the following Tab. 5.

For one trademark that is bound to at least one domain name, there is an average of 3.04

Tab. 4: Trademarks and domains found by using an algorithm of the exact match in words

The existence of a domain name in TM	Number of occurrences	%
The existence of a domain name for a TM	43,436	12.03%
There is no domain name for the TM	317,571	87.97%
Total	**361,007**	**100.00%**

Source: own

Tab. 5: Czech domain names and their binding to the trademarks

Domain name	Number of occurrences	%
Domain name is bound to a TM	123,050	21.65%
Domain name is not bound to a TM	445,222	78.35%
Total	**568,272**	**100.00%**

Source: own

domain names. One single registered domain name has 24,620 trademarks (56.68 %), two domain names have 6,716 trademarks (15.46%). 79.55% trademarks have the average number of 3 domain names per trademark. Although the average number of domain names to a single trademark is approx. 3 domain names, there are trademarks that have hundreds of them. As an example there are 239 domain names with a "Škoda" trademark, 110 with "Apple", 75 with "Bosh". For our analysis, it is important there are in average 63 domain names per trademark within the automotive industry, so it is possible to examine the usage of a trademark on such domains in more detail. It is also possible to ask a question whether an actual transcription of the trademark to a domain name follows some simple rules. For example, how often does a single-word trademark directly transcript to a domain name, e.g., "Škoda" to "skoda.cz", or multi-word "Potrefená husa" to "potrefenahusa.cz" or "potrefena-husa.cz". The frequency of these transcriptions is shown in Tab. 6. The biggest volume of direct transcription of a trademark into a domain name, such as "A" into "A.cz" amounts to 51.81% as expected. Multi-word transcriptions or those with an added character "-" are a rather minor issue (4.44%). The remaining 43.76% are other, more complex transcripts. Particularly those with an addition of

Tab. 6: The frequency of some transcriptions of trademarks to a domain name

Transcript (TM -> domain name)	Number of cases	%
A -> A.cz	31,851	51.81%
A B -> AB.cz	2,159	3.51%
A B C -> ABC.cz	152	0.25%
A B C D -> ABCD.cz	19	0.03%
A B -> A-B.cz	383	0.62%
A B C -> A-B-C.cz	13	0.02%
A B C D -> A-B-C-D.cz	2	0.00%
Another transcript (use of generic or descriptive words)	26,903	43.76%
Total	**61,482**	**100.00%**

Source: own

Tab. 7: The most common transcriptions of a trademark into a domain name (top 20)

Order	Pattern for TM	Number of domains
1	{TM}shop.cz	649
2	e{TM}.cz	538
3	i{TM}.cz	443
4	{TM}-shop.cz	334
5	{TM}club.cz	254
6	{TM}service.cz	245
7	auto{TM}.cz	224
8	{TM}group.cz	223
9	studio{TM}.cz	220
10	hotel{TM}.cz	217
11	{TM}web.cz	211
12	e-{TM}.cz	203
13	{TM}design.cz	200
14	{TM}praha.cz	199
15	for{TM}.cz	193
16	{TM}-praha.cz	181
17	{TM}brno.cz	170
18	{TM}plus.cz	156
19	{TM}reality.cz	153
20	salon{TM}.cz	151

Source: own

another generic or descriptive expression, such as a transcript of a trademark "A" to "prahaA.cz" or to "e-A.cz", etc.

The words, both generic and descriptive that are mostly used to form a domain name on the Czech Internet are "shop" (4,135 domains), "praha" (3,974 domains), "pro" (3,761 domains), "auto" (3,691 domains), "servis" (3,401 domains) and "brno" (3,220 domains). If, when using a division of a domain name into words, it is discovered that a trademark is incorporated into a domain name it is possible to examine patterns according to which a domain name is formed (beyond the simple transcription already referred to as "A" to "A.cz"). The most common patterns, where TM stands for a trademark, include TMshop.cz (649 domain names), eTM.cz (538 domain names), iTM.cz (443 domain names), TM-shop.cz (334 domain names), TMclub.cz (254 domain names), TMservis.cz (245 domain names), autoTM.cz (224 domain names), etc. The top 20 transcriptions are shown in Tab. 7.

3.1 Definition of Monitored Characters

In this section will be defined testable characters that will be detected for each trademark and its domain names using a computer program. The characters themselves originate in existing decisions of the WIPO and CAC directly. Some characters were derived indirectly so their influence on trademark infringement is to be further verified. Characters will be further analyzed through regression analysis and we will ask a question of how the existence of found characters relates to authorized/unauthorized use of a trademark on a given domain/website.

For the purpose of this article, 17 characters were defined to be further examined among domain names/websites associated with trademarks within the automotive industry. In the following text, we will use an expression "website has a given character" if the conditions defining the character are met (e.g., the website contains HTML frames element). As shown further, some of the characters are either not

much used in practice (small representation across domain names examinees) or they are used extensively, but do not have a relevant value in order to determine whether the occurrence of the character rather leads to trademark infringement or not. The characters have been defined on the basis of an earlier analysis on indicators of domain names in conflict-of-law (Gongol, 2013). Here are the characters in question:

00 – No Page - Website has no content;
01 – Park - Website is "parked";
02 – Forward - Website is automatically redirected to another website;
03 – Size - Website content is very small;
04 – GLinks - Website contains a link to an official website of a trademark owner;
05 – Title - Website name contains a text of a trademark;
06 – GKeywords - Website's metadata contain a text of a trademark;
07 – SKeywords - Website's metadata contain suspicious words;
08 – Ads - Website contains advertising;
09 – SURL - Website's URL is suspicious;
10 – Frames - Website contains HTML frames;
11 – SContent - Website of suspicious content;
12 – GOwner - Domain name belongs to a trusted holder;
13 – BForward - Website has an automatic redirection to the competition;
14 – BOwner - Website is located on a domain belonging to a suspicious holder;
15 – BLinks - Website contains a link to a website in conflict-of-law;
16 – NoTM - Website does not contain a reference to a trademark.

Tab. 8: Trademarks with the largest percentage of domains categorized as *in collision*

Trademark	In collision	%	Number of domains
kia	21	55.26%	38
mercedes	30	54.55%	55
renault	46	52.27%	88
toyota	30	49.18%	61
peugeot	26	49.06%	53
mazda	17	45.95%	37
fiat	33	43.42%	76
ford	56	42.75%	131
audi	17	42.50%	40
skoda	96	41.03%	234
hyundai	29	40.28%	72
bmw	44	40.00%	110
jeep	12	37.50%	32
mitsubishi	9	36.00%	25
suzuki	16	35.56%	45
opel	23	35.38%	65
porsche	14	34.15%	41
honda	21	31.82%	66
nissan	12	31.58%	38
subaru	13	30.95%	42
citroen	25	30.12%	83

Source: own

3.2 Manual Categorization of Domain Names that are Relevant to the Sector of the Car Industry

Bearing in mind the objective of the article, detailed analysis is focused on the sector of the automotive industry particularly because in this sector exists in average a large number of relevant domain names to trademarks (63 domain names to a single trademark). Other reasons are the anticipated volume of investments in this sector and also a general awareness and demand from ordinary Internet users. Noteworthy is also the fact that the trademark in the automotive sector has a strong distinctive character, which reduces the risk of interchangeability.

All trademarks for which there was at least one relevant Czech domain name were selected for the analysis. If, in addition to the car manufacturer's trademark there were domain names for a particular type of product linked with the trademark (such as "Fabia"), these domain names were also included in the selection. Using an algorithm of the exact match at the level of words there were found 1,825 Czech domain names to such trademarks.

The intention of manual categorization process is to go through these domain names and determine, whether they are in conflict-of-law*/in collision* - if so, it is a case of trademark infringement or they are in no conflict-of-law/*non-collision* - in which case there is no unauthorized use of a trademark (a website is the official website of the owner of the trademark or it has been used legitimately by a third party). Of a total number of 1,621 existing domain names, there were 664 (40.96%) categorized as *collision*, the remaining 957 (59.04%) as a *non-collision*. For example, the trademark *Škoda* has 96 domain names *in collision* out of 234 relevant ones (41.03%), the *BMW* trademark has 44 *in collision* out of 110 (40.00%). Among trademarks which have the largest number of domain names categorized as *in collision* are *Kia* (55.26%), *Mercedes* (54.55%), *Renault* (52.27%), for more see Tab. 8.

First, while examining domain names in detail, it is important to find repetitive patterns that were used when forming a domain name out of a trademark. Once it is done it is possible to determine how many times the patterns were used for domain names that are categorized as *in collision* and *non-collision*. For example, the pattern *TMweb.cz*, from which may arise domain names such as *skodaweb.cz*, *fordweb.cz*, *renaultweb.cz*, etc. were used in 28 cases, all of which have been categorized as domains *in collision*. Similarly, the *chiptuning-TM.cz* pattern was used in 16 cases and in all of them the domain was categorized as *in collision*. For a detailed overview of the top twenty of the most frequent patterns, see Tab. 9.

3.3 Regression Analysis of the Relationship between Characters and the Category of a Eomains Name

In this part, we will focus on connections between the occurrence of specific characters and assigning of a domain name into *collision* or *non-collision* category. The aim is to identify a set of characters for which the automatic tracking makes the most sense and which determine the domain's probable category with an acceptable level of error. Then it would be possible to extend the results outside the area of the automotive industry in the follow-up research and use them for the next automatic analysis of domain names on the Internet. However, as mentioned above, it will not be possible to generalize the statistical findings to all domain names, as the result would be negatively influenced by a selective error. For all Czech domain names that were closely examined by the computer program were found their specific characters (total of 1,621 domain names). The three most commonly represented characters are *05 – Title* (66,66% of domain names), *06 – GKeywords* (47,13% of domain names), and *11 – SContent* (25,35% of domain names). On the other hand, the characters *13 – BForward* and *10 – Frames* do not appear in examined domain names almost at all (occurrence under 1%). Therefore, given their insignificance, they will not be included in further analysis. The scale of the occurrence of individual characters is shown in Tab. 10 in detail.

Before a logistic regression model is applied, we will examine the relationship between the occurrence/absence of a specific character and a resulting category of a domain name applying the method of conditional probability.

$$\mathrm{P}(A \mid B) = \frac{P\,(A \cap B)}{P\,(B)} \qquad (1)$$

Tab. 9: **Top 20 patterns found in studied domains and their amount of representation in "collision" category**

Pattern in domain	Number of domains	In collision	In collision %	Non-collision	Non-collision %
{TM}web.cz	28	28	100.00%	0	0.00%
{TM}club.cz	27	18	66.67%	9	33.33%
{TM}-auto.cz	20	4	20.00%	16	80.00%
chiptuning-{TM}.cz	16	16	100.00%	0	0.00%
{TM}-klub.cz	15	12	80.00%	3	20.00%
{TM}-forum.cz	15	15	100.00%	0	0.00%
{TM}-club.cz	13	5	38.46%	8	61.54%
{TM}-praha.cz	12	7	58.33%	5	41.67%
{TM}levne.cz	12	12	100.00%	0	0.00%
{TM}praha.cz	11	4	36.36%	7	63.64%
portal-{TM}.cz	11	11	100.00%	0	0.00%
{TM}centrum.cz	11	7	63.64%	4	36.36%
{TM}brno.cz	10	3	30.00%	7	70.00%
{TM}dily.cz	10	6	60.00%	4	40.00%
{TM}klub.cz	10	6	60.00%	4	40.00%
{TM}-servis.cz	9	5	55.56%	4	44.44%
{TM}-dily.cz	9	2	22.22%	7	77.78%
nahradnidily{TM}.cz	8	8	100.00%	0	0.00%
autovrakoviste{TM}.cz	8	6	75.00%	2	25.00%
{TM}vrakoviste.cz	8	5	62.50%	3	37.50%

Source: own

The statement $P(A \mid B)$ expresses the conditional probability of the phenomenon A provided that there occurred phenomenon B (e.g, the probability of the *collision* provided there occurred the character *"with 08"*). For example, to determine the probability of categorizing a domain name as *in collision* ($P(A \mid B)$) provided the occurrence of character B it is necessary to know the probability of its occurrence in domain names in *collision* ($P(A \cap B)$) and divide it by the overall probability of the occurrence of character B across all examined domains ($P(B)$). Tab. 11 contains a list of probabilities for all examined characters that we can use to calculate all conditional probabilities.

Further on (in Tab. 12) there are calculated conditional probabilities in all combinations of occurrence/absence of characters. These lead to a determination of category *in collision* (it is then possible to calculate the probability of *non-collision* by counting the remainder to 100%). Lines in red relate to characters for which there is a low percentage of occurrence (under 1%) and which, despite high probabilities, will not be considered given their inconclusiveness (e.g, *13 – BForward* 100%). From Tab. 12, it is also evident that the absence of a particular character (column *Without Character*) does not have a significant impact on the determination of category since the data is around 50% (thus in the range of coincidence, given there are only two possible options - *collision/non-collision*).

In the case of characters *00 – NoPage, 01 – Park, 08 – Ads, 09 – SURL, 11 – SContent, 14 – BOwner* and *16 – NoTM* it is obvious they are more likely to lead to the *collision* category. The characters *04 – GLinks* and *12 – GOwner* to the category of *non-collision*. For the following logistic regression (besides characters 00 to 16),

Tab. 10: Detailed scale of occurrence of individual characters across domain names of the automotive industry

Identification of characters	With character	%	Without character	%
05 – Title	1,080	66.63%	541	33.37%
06 – GKeywords	764	47.13%	857	52.87%
11 – SContent	411	25.35%	1,210	74.65%
02 – Forward	329	20.30%	1,292	79.70%
04 – GLinks	306	18.88%	1,315	81.12%
03 – Size	303	18.69%	1,318	81.31%
14 – BOwner	289	17.83%	1,332	82.17%
12 – GOwner	285	17.58%	1,336	82.42%
15 – BLinks	237	14.62%	1,384	85.38%
08 – Ads	203	12.52%	1,418	87.48%
16 – NoOz	132	8.14%	1,489	91.86%
01 – Park	111	6.85%	1,510	93.15%
09 – SURL	51	3.15%	1,570	96.85%
00 – NoPage	39	2.41%	1,582	97.59%
07 – SKeywords	22	1.36%	1,599	98.64%
13 – BForward	9	0.56%	1,612	99.44%
10 – Frames	7	0.43%	1,614	99.57%

Source: own

Tab. 11: Table of probabilities of occurrences of tracked characters

Identification of the character	With character & in collision	With character & non-collision	Without character & in collision	Without character & non-collision
00 – NoPage	34	5	630	952
01 – Park	106	5	558	952
02 – Forward	98	231	566	726
03 – Size	153	150	511	807
04 – GLinks	18	288	646	669
05 – Title	368	712	296	245
06 – GKeywords	286	478	378	479
07 – SKeywords	20	2	644	955
08 – Ads	200	3	464	954
09 – SURL	42	9	622	948
10 – Frames	1	6	663	951
11 – SContent	355	56	309	901
12 – GOwner	18	267	646	690
13 – BForward	9	0	655	957
14 – BOwner	262	27	402	930
15 – BLinks	177	60	487	897
16 – NoOz	108	24	556	933

Source: own

Tab. 12: The conditional probability of occurrence/absence of characters and their effect on the probability of determining the *collision/non-collision* category

Character - factor	With character & in collision	Without character & in collision
00 – NoPage	87.18%	39.82%
01 – Park	95.50%	36.95%
02 – Forward	29.79%	43.81%
03 – Size	50.50%	38.77%
04 – GLinks	5.88%	49.13%
05 – Title	34.07%	54.71%
06 – GKeywords	37.43%	44.11%
07 – SKeywords	90.91%	40.28%
08 – Ads	98.52%	32.72%
09 – SURL	82.35%	39.62%
10 – Frames	14.29%	41.08%
11 – SContent	86.37%	25.54%
12 – GOwner	6.32%	48.35%
13 – BForward	100.00%	40.63%
14 – BOwner	90.66%	30.18%
15 – BLinks	74.68%	35.19%
16 – NoOz	81.82%	37.34%

Source: own

will be considered another categorical variable to determine whether a domain name is in *collision* or not (so-called *Collision Factor*). In order to find relationships between characters and the Collision Factor will be applied a *logistic regression* with the level of significance α = 5%. When examining the dependencies it is necessary to prevent a strong correlation links, which could negatively affect the resulting regression model. This way it is possible to find 16 statistically significant relationships out of 153 possible. As expected, the Collision Factor is strongly correlated (> 0.33) to characters *04 – GLinks*, *08 – Ads*, *11 – SContent*, *12 – GOwner*, *14 – BOwner*. Also, there are dependencies between the individual characters: strongly correlated (> 0.4) is a *16 – NoTM* with characters *01 – Park, 03 – Size, 05 – Title*, then character *14 – BOwner* with *08 – Ads, 11 – SContent* and character *15 – BLinks* with *11 – SContent* and *14 – BOwner*. We take these dependencies into account when reducing the found logistic model. The remaining dependencies have a coefficient below 0.3 so these rates of association can be considered insignificant.

It is now possible to create a logistic regression model (see, for example, Kleinbaum et al., 2010) in order to describe the association between both explanatory factors, "characters" and dependent "Collision Factor". As mentioned, for the lack of occurrence the characters *10 – Frames* and *13 – BForward* are excluded from the default model. For the sake of brevity, [00] [01] [02] in square brackets reflect the categorical parameters of independent characters *00 – NoPage, 01 – Park, 02 – Forward*, etc.

$$\begin{aligned} &logit(P(collision=1)) \\ &= \beta_0 + \beta_1 \times [00] + \beta_2 \times [01] + \beta_3 \times [02] + \\ &+ \beta_4 \times [03] + \beta_5 \times [04] + \beta_6 \times [05] + \\ &+ \beta_7 \times [06] + \beta_8 \times [07] + \beta_9 \times [08] + \\ &+ \beta_{10} \times [09] + \beta_{11} \times [11] + \beta_{12} \times [12] + \\ &+ \beta_{13} \times [14] + \beta_{14} \times [15] + \beta_{15} \times [16] \end{aligned} \quad (2)$$

Model converged 14 iterations from the default zero model with a value of -2 Log (L) = 2,193.9304 to the target value -2 Log (L) = 731.4205. The values of all the coefficients found including the limits (α = 0.05) for each factor are shown in Tab. 13.

Tab. 13: Default logistic model

Beta	Character	Coefficient	Deviation	p	E.R.	Min (95%)	Max (95%)
1	00 – NoPage	4.1751	0.5424	0	65.0472	22.4668	188.3283
2	01 – Park	4.6121	0.6556	0	100.691	27.856	363.9668
3	02 – Forward	0.4003	0.2603	0.1242	1.4922	0.8958	2.4857
4	03 – Size	1.1558	0.2607	0	3.1767	1.9058	5.295
5	04 – GLinks	-1.2359	0.3639	0.0007	0.2906	0.1424	0.5929
6	05 – Title	-0.7502	0.2493	0.0026	0.4723	0.2897	0.7697
7	06 – GKeywords	-0.0494	0.2283	0.8286	0.9518	0.6084	1.4889
8	07 – SKeywords	-1.432	1.0313	0.165	0.2388	0.0316	1.8029
9	08 – Ads	5.6256	0.6318	0	277.4512	80.432	957.0709
10	09 – SURL	0.4696	0.7433	0.5275	1.5994	0.3726	6.8648
11	11 – SContent	4.0475	0.2539	0	57.2561	34.8088	94.1792
12	12 – GOwner	-2.0854	0.3728	0	0.1243	0.0598	0.258
13	14 – BOwner	2.3017	0.3436	0	9.9915	5.0949	19.594
14	15 – BLinks	0.68	0.3801	0.0736	1.9739	0.937	4.1579
15	16 – NoTM	1.7874	0.3767	0	5.9736	2.855	12.4986
0	Intercept	-2.3188	0.2509	0			

Source: own

Tab. 14: Reduced logistic model

Beta	Character	Coefficient	Std. Deviation	p	E.R.	Min (95%)	Max (95%)
1	00 – NoPage	4.0727	0.5307	0	58.7168	20.7511	166.1438
2	01 – Park	4.1072	0.5209	0	60.7741	21.893	168.7069
3	03 – Size	1.1444	0.2558	0	3.1404	1.9021	5.185
4	04 – GLinks	-1.1627	0.3603	0.0013	0.3126	0.1543	0.6335
5	05 – Title	-0.7749	0.2395	0.0012	0.4608	0.2882	0.7367
6	08 – Ads	5.7617	0.6316	0	317.8989	92.1941	1096.1626
7	11 – SContent	4.1125	0.2491	0	61.1005	37.4966	99.5629
8	12 – GOwner	-2.0815	0.3691	0	0.1247	0.0605	0.2572
9	14 – BOwner	2.3286	0.3298	0	10.2639	5.3771	19.592
10	16 – NoTM	1.8956	0.3676	0	6.6564	3.2387	13.6805
0	Intercept	-2.2164	0.2245	0			

Source: own

Tab. 15: The test of credibility (likelihood) of the zero, default and reduced model

Logistic model	#df	-2 Log (L)	df	D (chi-quadrate)	P (>D)
Reduced model	11	739.1572			
Zero model	1	2,193.9304	-10	1,454.7732	<0.00001
Default model	16	731.4205			
Reduced model	11	739.1572	-5	7.7367	0.1713

Source: own

The coefficients for the characters [02], [06] [07], [9] and [15] have p-values greater than set significance level. Therefore, they are excluded from the reduced model. The original model, which contained 15 factors, is being reduced to a model with 10 factors having an equation:

$$\begin{aligned} &logit(P(collision=1)) \\ &= \beta_0 + \beta_1 \times [00] + \beta_2 \times [01] + \beta_3 \times [03] + \\ &+ \beta_4 \times [04] + \beta_5 \times [05] + \beta_6 \times [08] + \\ &+ \beta_7 \times [11] + \beta_8 \times [12] + \beta_9 \times [14] + \\ &+ \beta_{10} \times [16] \end{aligned} \quad (3)$$

Reduced model converged 14 iterations to the value -2 Log (L) = 739.1572. The calculated coefficients, including the limits of the estimate of the relative risk (E.R.), are shown in the following Tab. 14.

We will compare the significance of the reduced model with the zero one and also the default model by using the test of credibility (likelihood). The reduced model is significantly different from the zero one ($p < 0.00001$) and it is not remarkably different from the default model ($p = 0.1713$), see Tab. 15.

Now we will create separate logistic models, in which we will examine the influence of every single character on the Collision Factor. Comparison of the relative change in the coefficients of a standalone logistic models with coefficients of the relative model can help reveal hidden dependencies between the characters. Relative changes in the coefficients are shown in Tab. 16.

It is clear that for the character *03 – Size* there is a multiple value change (from 0.4768 to 1.1144), which is explainable by a strong association with *01 – Park* (Cramer's coefficient is 0.352) and character *16 – NoTM* (Cramer's coefficient is 0.487). From the reduced model we will create the final model from which we will exclude the character *03 – Size*. Also the character *05 – Title* is strongly associated with the other parameters of the model (with

Tab. 16: Relative changes in the coefficients of standalone models and the reduced model

Parameter	Standalone model	Reduced model	Relative change
00 – NoPage	2.3298	4.0727	-0.7481
01 – Park	3.5882	4.1072	-0.1446
03 – Size	0.4768	1.1444	-1.4002
04 – GLinks	-2.7376	-1.1627	0.5753
05 – Title	-0.8491	-0.7749	0.0874
08 – Ads	4.9205	5.7617	-0.1710
11 – SContent	2.9169	4.1125	-0.4099
12 – GOwner	-2.6310	-2.0815	0.2089
14 – BOwner	3.1112	2.3286	0.2515
16 – NoTM	2.0217	1.8956	0.0624

Source: own

Tab. 17: Test of credibility of the reduced and final model

Logistic model	#df	-2 Log (L)	df	D(chi-quadrate)	P (>D)
Reduced model	11	739.1572			
Final model	9	769.0925	-2	29.9353	<0.00001
Final model	9	769.0925			
Zero model	1	2,193.9304	-8	1,424.8379	<0.00001

Source: own

the character *01 – Park* - Cramer's coefficient 0.331) and the character *16 – NoTM* - Cramer's coefficient 0.42), therefore we will not consider it in the final model. The final model converged k -2Log (L) = 769.0925 after 14 iterations. When comparing the reduced and final model using the credibility test, the final model differs significantly from the reduced one, see Tab. 17.

We can write the equation of the resulting model as follows:

$$ln\left(\frac{P(collision=1)}{1-P(collision=1)}\right) = -2.3284 + 4.1848 \times [00] + 4.4662 \times \times [01] - 1.4546 \times [04] + 5.3405 \times \times [08] + 3.7252 \times [11] - 2.0853 \times \times [12] + 2.1862 \times [14] + 2.9273 \times [16] \quad (4)$$

where [00] is 1 if the domain has the given character and value of 0 if the domain name does not have the given character (by analogy to [01] [04], ...). The list of found coefficients including p-value is shown in Tab. 18.

From this can be concluded that the existence of the character *08 – Ads* significantly increases the chance that a website is in *collision* (about 200 times, column E.R. = odds ratio of the *collision/non-collision*). Similarly, character *00 – NoPage* increases the chance of it being in *collision* approximately 65 times, character *01 – Park* 87 times. On the other hand, the existence of character *12 – GOwner* reduces the chance of it being in *collision* 8 times. Similarly, the character *04 – GLinks* reduces the chance of about 5 times. The resulting model equation can be used to estimate the probability of "being in collision" based on the information of found characters for a given domain. For example, suppose a website has two characters: 08 – Ads and 11 – SContent. After substituting, the equation will look like this:

$$ln\left(\frac{P(collision=1)}{1-P(collision=1)}\right) = -2.3284 + 4.1848 \times 0 + 4.4662 \times \times 0 - 1.4546 \times 0 + 5.3405 \times 1 + + 3.7252 \times 1 - 2.0853 \times 0 + 2.1862 \times \times 0 + 2.9273 \times 0 = 6.737 \quad (5)$$

Tab. 18: The final logistic model

Beta	Character	Coefficient	Deviation	p	E.R.	Min (95%)	Max (95%)
1	00 – NoPage	4.1848	0.5028	0	65.6785	24.5148	175.9616
2	01 – Park	4.4662	0.5107	0	87.0272	31.9864	236.7799
3	04 – GLinks	-1.4546	0.3433	0	0.2335	0.1191	0.4576
4	08 – Ads	5.3405	0.6233	0	208.6105	61.4928	707.6978
5	11 – SContent	3.7252	0.2223	0	41.4796	26.8271	64.1349
6	12 – GOwner	-2.0853	0.3619	0	0.1243	0.0611	0.2526
7	14 – BOwner	2.1862	0.3086	0	8.9017	4.8613	16.3002
8	16 – NoTM	2.9273	0.3215	0	18.6778	9.9454	35.0775
0	Intercept	-2.3284	0.1469	0			

Source: own

After removing the logarithm and the expression of values P (collision = 1), we get:

$$P(collision = 1) = \frac{e^{6.7373}}{1 + e^{6.733}} = 0.9988 = 99{,}9\% \quad (6)$$

Therefore, if the page has the characters 08 and 11, then there is 99.9% probability that according to the model the website is in collision.

Discussion

When determining the relevance of a domain name for a particular trademark, an algorithm was used that sought the best breakdown of the domain name into individual words with the help of Czech and English language dictionaries and other support lists by using a function in its maximum potential. In the light of current technologies, it seems best for this type of task to use a self-taught neural or convolutional neural network that would itself find the best divisions.

Another part was aimed at finding the relevant characters, which typically occur on trademark infringing websites. The analysis showed, which characters should be followed further and which are of lower significance. In practical terms, there is a difficulty in the algorithmic implementation of characters for which it is necessary to know the specifics of the trademark sector. In the case of the examined automotive industry, it is a network of group links between automobile manufacturers, types of cars and their links, words whose appearance on the website is suspicious from the perspective of the automotive industry (for example, the word “perfume”, “accommodation”, etc.). Generalization of the analysis across other sectors, or the “whole Internet” in domain .cz, would be intriguing and under certain conditions even possible. Some characters that showed high relevance in determining domain name being in collisions do not require a sectoral knowledge (for example, character 08 – Website contains advertising, character 00 – NoPage or character 01 – Park) – an extension of the analysis would be straightforward for this type of characters. For some characters, such as character 11 – Website of suspicious content, a sectoral knowledge is necessary. However, obtaining such knowledge could also be possible through a well-designed self-taught algorithm based on neural networks. The analysis was in many ways based on a connection between an existing lexical trademark and a textual binding to the respective domain names (word division). However, there could be other binding options - for example, a phonetic one, or an option that takes typing errors into account or ignores the Czech diacritics.

Holders of domain names who commit a trademark infringement always do so for the sole purpose of their own benefit. They need to have Internet users navigated to their websites when searching a specific trademark. In some cases, there are holders controlling domain names in collision only, e.g. JAKUB-ELIAS controls 32 registered domain names with 100% of them being in collision and so do the others, e.g. WEBDEVEL, PROFIWH-PETRANOVAKOVA, MITONCZ. An Internet user who needs to find relevant websites of a trademark uses either a direct URL, which they deduce by an intuitive transcription of the trademark into a domain name and put it in the address bar of the browser or they use a search portal. The analysis showed the relevance of direct transcription of a trademark into a domain name in 51.81% of registered trademarks. A common SEO practice to embed keywords and relevant information into a domain name (in our case, a trademark) retreats with the advent of more advanced search algorithms. Therefore, it will be beneficial to link the user, trademark and relevant domain names via Internet search engines themselves, i.e. enter the trademark into the search engine using a program and identify top 10 results as relevant.

Detailed analysis dealt with the influence of characters on the dichotomy factor of collision/non-collision, mainly in order to find characters that typically lead to a domain name in collision and that could be used, for example, in a computer program for automatic search of websites in collision. For practical purposes, in any other analysis, it will be useful to divide the collision factor into multiple categories according to the severity or nature of the infringement (systematic trademark infringement, misuse of a trademark arising from lack of knowledge, etc.).

Conclusions

The text analyzed the links of 568 thousand Czech domain names in relation to 361

thousand trademarks from IPO registry. For the 43 thousand trademarks (12%) there exists at least one relevant Czech domain name and 123 thousand domain names (21.65%) have a connection to a trademark. There is an average of 3 domain names per trademark that has at least one domain name. In the sector of the automotive industry that was examined in detail, there are even 63 domain names per trademark. Over 51% of trademarks use a direct transcription of the trademark name into the domain name by excluding gaps and removing diacritics. Other most common transcriptions of trademarks include (TM)shop.cz, e(TM).cz, i(TM).cz, (TM)-shop.cz, (TM)club.cz. In accordance with research questions, the analysis focused on the automotive sector, where there were examined 1621 domain names in detail, of which 664 (40.96%) were categorized as in *collision* because of the found trademark infringement. The following trademarks show the highest percentage of website abuse: Kia (55.26%), Mercedes (54.55%), Renault (52.27%) and Toyota (49.18%). The lowest percentage applies to Citroen (30.12%), Subaru (30.95%), Nissan (31.58%) and Honda (31.82%). Patterns typical for trademark infringement within automotive industry include (TM)web.cz, (TM)levne.cz, chiptuning-(TM).cz, portal-(TM).cz. The first research question, whether trademark infringement is a marginal issue or widespread practice, can be answered, at least in the case of the automotive industry, in favor of a widespread practice.

The second research question concerning the existence of similar or mutual characters of domain names/websites in violation of trademark rights, we need to answer in the affirmative. Based on the analysis of the domain disputes, there were examined 17 characters and their effect on increasing the collision risk of a domain name. The following characters are among the significant factors: 08 – Website contains advertising, 11 – Website of suspicious content, 16 – Website does not contain a reference to a trademark, 00 – Website has no content, 01 – Website is "parked", 14 – Website is located on a domain belonging to a suspicious holder. Conversely, the presence of the following characters significantly reduces the risk of a domain name being in collision: 04 – Website contains a link to an official website of a trademark owner and character 12 – Domain name belongs to a trusted holder. According to the analysis, some of the other characters statistically proved to be insignificant (e.g. 02 – Website is automatically redirected to another website).

The last research question tackled the possibility of an automated search for specific cases of unauthorized use of trademarks, resp. finding general rules for such a system. Partial analyses were focused on the automotive industry. A simple extension over the entire Internet, resp. all domain names would be burdened with a selection error. Nevertheless, generalizing the analysis across other sectors or the entire Internet within ".cz" domain is possible under certain conditions. Some characters that reported high relevance in determining domain name collisions do not require a sectoral knowledge (e.g., 08 – Website contains advertising, 00 – Website has no content, 01 – Website is "parked"/has unreleased content) – for these characters the extension of the analysis is straightforward. For some others, such as character 11 – Website of suspicious content, the sectoral knowledge is necessary (this can be obtained, for example by a well-designed self-taught algorithm based on neural networks).

The above also shows the possible direction for further research. In addition to the extension of the analysis across the entire ".cz" domain it is possible, within the framework of the methodology to keep adding new characters that indicate a domain name in collision/non-collision. For example, the links formed by the phonetic similarities or typos, date of registration of the trademark against domain name registration, prevention of access of indexing robots to Websites, etc. Finding a link between a trademark and a domain name based on Internet search engines would also be highly beneficial. Not only for the purpose of identifying the trademark's distinctive character, but also to find relevant domain names that appear in the references as natural search results.

The results of the carried out research show practical possibilities to limit the trademark infringement on the Internet by automated means that can be used by both, the entities protecting the rights of trademark proprietors and, where appropriate, administrators of national or generic domains, provided they accepted the possibility, albeit only of a partial,

automated process of detecting a registered domain name collision with an existing trademark.

This paper was supported by the Student grant competition project SGS/7/2017: "Acceptance of technology from the perspective of marketing tools."

References

Bettinger, T., Willoughby, T., & Abel, S. (2005). *Domain name law and practice: an international handbook*. New York, NY: Oxford University Press.

Burmann, C., Riley, N. M., Halaszovich, T., & Schade, M. (2017). *Identity-Based Brand Management*. Wiesbaden: Springer Gabler.

Branthover, N. (2002). UDRP – A Success Story: A Rebuttal to the Analysis and Conclusions of Professor Milton Mueller in „Rough Justice". International Trademark Association. Retrieved July 23, 2019, from https://www.inta.org/Advocacy/Documents/INTAUDRPSuccesscontraMueller.pdf.

Crass, D., Czarnitzki, D., & Toole, A. A. (2019). The Dynamic Relationship Between Investments in Brand Equity and Firm Profitability: Evidence Using Trademark Registrations. *International Journal of the Economics of Business*, *26*(1), 157-176. https://doi.org/10.1080/13571516.2019.1553292.

Edwards, L., & Waelde, Ch. (2009). *Law and the internet*. 3rd ed. Oxford: Hart publishing.

Gielen, Ch. (2010). Keyword advertising and European Trade Mark Law, Retrieved July 22, 2019, from http://charlesgielen.com/2.html.

Gongol, T. (2012). The Analysis of Czech Arbitral Court Verdicts in Cases. *DANUBE/Law and Economics Review, 3*(1), 71-93.

Gongol, T. (2013). *Právní aspekty nekalé soutěže na internetu.* Karviná: Silesian University in Opava, Faculty of Business Administration in Karviná.

Gongol, T. (2013b). The Preliminary Ruling Decision in the Case of Google vs. Louis Vuitton Concerning the AdWord Service and its Impact on the Community Law. *The Amfiteatru Economic Journal, 15*(33), 246-260.

Gongol, T. (2014). Contribution to the discussion on the notion of 'Alternative dispute resolutions'. *European Offroads of Social Science, 5*(1), 3-15.

Gongol, T. (2016). Judicial practice in selected countries in cases of legal liability of auction portals for the sale of counterfeits. *Acta Academica Karviniensia, 16*(4), 23-31. https://doi.org/10.25142/aak.2016.029.

Griffiths, A. (2008). A Law-and-Economic Perspective on Trade Marks. In: L. Bently, J. Davis, J. C. Ginsburg (ed). *An Interdisciplinary Critique* (pp. 241-267). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511495212.012.

Halvorson, T., Der, M. F., Foster, I., Savage, S., Saul, L. K., & Voelker, G. M. (2015). From .Academy to .Zone: An Analysis of the New TLD Land Rush, In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference,* (pp. 381-394). San Diego, CA: University of California. https://doi.org/10.1145/2815675.2815696.

Horáček, R., Čada, K., & Hajn, P. (2017). *Práva k průmyslovému vlastnictví. 3rd ed.* Praha: C. H. Beck.

Janis, M., & Dinwoodie, G. (2007). Confusion Over Use: Contextualism in Trademark Law. *Iowa Law Review, 92,* 1597-1667.

Jansa, L., Otevřel, P., Čermák, J., Mališ, P., Hostaš, P., Matějka M., & Matejka, J. (2016). *Internetové právo*. Brno: Computer Press.

Janouch, V. (2014). Internetový marketing. 2nd ed. Brno: Computer Press.

Kleinbaum, D., Klein, M., & Pryor, E., R. (2010). Logistic regression: a self-learning text. 3rd ed. New York, NY: Springer.

Klerman, D. (2016). Forum Selling and Domain-Name Disputes. *Loyola University Chicago Law Journal, 48*(2), 561–584.

Korczynski, M., Wullink, M., Tajalizadehkhoob, S., Moura, G. C., & Hesselman, C. (2017). Statistical Analysis of DNS Abuse in gTLDs. ICANN. Retrieved July 23, 2019, from https://www.icann.org/en/system/files/files/sadag-final-09aug17-en.pdf.

Lloyd, I. J. (2011). *Information Technology Law.* 6th ed. New York, NY: Oxford University Press.

Lukose, L. (2013). Consumer Protection vis a vis Trademark Law. *International Journal on Consumer Law and Practise, 1*(2013), 89–101.

Merges, R., Menell, P. S., & Lemley M. A. (2012). *Intellectual property in the new technological age.* 6th ed. New York, NY: Wolters Kluwer.

Munková, J., Kindl, J., & Svoboda, P. (2012). *Soutěžní právo. 2nd ed.* Prague: C. H. Beck.

Oullette, L. (2014). The Google Shortcut to Trademark Law. *California Law Review, 102*(2), 351-407. https://doi.org/10.2139/ssrn.2195989.

Otim, S., & Grover, V. (2010). E-commerce: a brand name's curse. *Electronic Markets, 20*(2), 147-160. https://doi.org/10.1007/s12525-010-0039-6.

Pelikánová, R. (2012). *Ekonomické, právní a technické aspekty doménových jmen v globální perspektivě*. Ostrava: Key Publishing.

Polčák, R. et al. (2018). Právo informačních technologií. Praha: Wolters Kluwer.

Saunders, K. M., & Berger-Walliser, G. (2011). The Liability of Online Markets for Counterfeit Goods: A Comparative Analysis of Secondary Trademark Infringement in the United States and Europe. *Northwestern Journal of International Law & Business, 32*(1), 37-92.

Senftleben, M. R. F. (2012). Keyword Advertising in Europe – How the Internet Challenges Recent Expansion of EU Trademark Protection. *Connecticut Journal of International Law, 2012*(27), 39-74.

Slováková, Z. (2006). *Průmyslové vlastnictví.* 2nd ed. Prague: LexisNexis.

Vissers, T., Wouter, J., & Nikiforakis, N. (2015). Parking Sensors: Analyzing and Detecting Parked Domains. In *Network and Distributed System Security Symposium*. San Diego, CA: Internet Society. https://doi.org/10.14722/ndss.2015.23053.

Werra, J. (2016). Alternative Dispute Resolution in Cyberspace: The Need to Adopt Global ADR Mechanisms for Addressing the Challenges of Massive Online Micro-Justice. *Swiss Review of International and European Law, 26*(2), 289–306. https://doi.org/10.2139/ssrn.2783213.

doc. Mgr. Tomáš Gongol, Ph.D.
Silesian University in Opava
School of Business Administration in Karvina
Department of Business Economics
and Management
Czech Republic
tomas.gongol@slu.cz

Abstract

TRADEMARK INFRINGEMENTS IN THE DOMAIN ".CZ"

Tomáš Gongol

The aim of this article is to fill a gap in an area that has not yet been closely examined in the Czech Republic and the world: examining the level of trademark infringement in relation to the basic elements of the logical architecture of the Internet, namely domain names, resp. Websites that are published on them. The article aims to determine the actual state and create a methodology of rapid, and to some extent automated detection of the collision of rights connected to trademarks with domain names. Not only for the large scope of the investigated subject has it focused on the sector of the automotive industry in particular. Given the aim of the work, it answers questions whether the phenomenon of trademark infringement on the Internet is rather a minor issue, which concerns only a fraction of domain names and websites; whether the domains names, resp. websites violating the trademark rights share any similar characters; whether it is possible to automate the search process of finding trademark infringements or if there is a way of finding general rules that can be a valuable help in this process. The first part of the article describes the used methodology, sources of available data and the way the data were worked with. The next part deals in detail with input data that are relevant for this article. It describes the ways the data were obtained and the constraints that needed to be overcome doing so. Basic statistical parameters of the input data are also mentioned. The third part is focused on the important findings found in input data relating to Czech domain names and trademarks used on the Internet followed by detail examination of the Czech domain names for the sector of the automotive industry. By using the defined indicators of a collision (the characters) the results of the article show that the domain names on which the trademark infringement has been committed, share the same set of characters that can be tested automatically by a computer program.

Keywords: *Alternative dispute resolution, domain name, logistic regression, trademark infringement.*

JEL Classification: *K24.*

DOI: *10.15240/tul/001/2019-4-011.*