Ako čítať a interpretovať „Protokol o kontrole originality“

(pre protokoly systému CRZP/ANTIPLAG)

Aktualizácia: 20.3.2016

Protokol o kontrole originality (ďalej len protokol) je výstupom zo „Systému na odhaľovanie plagiátov pre potreby slovenských akademických a výskumných inštitúcií“ (ďalej ANTIPLAG), ktorý beží nad Centrálnym registrom záverečných a kvalifikačných prác (ďalej CRZP). Protokol je elektronický dokument generovaný algoritmom na odhaľovanie plagiátov. Úlohou tohto dokumentu je vysvetliť jednotlivé časti protokolu.

POZOR! V závislosti na rozsahu práce a množstve nájdenej zhody môže mať v niektorých prípadoch protokol pomerne značný rozsah. Pred vytlačením protokolu zvážte rozsah tlače. Pre účely identifikácie protokolu a získania základnej informácie o výsledku je podstatná prvá strana protokolu.

Obsah

Všeobecné údaje protokolu 2
Záhlavie protokolu 2
Grafické zvýraznenie pokrytia textu 2
Čistý text (plain text) - informácie 2
Početnosť slov – histogram 3
Zoznam dokumentov 3
Podobný text – podobný odsek 3
Obmedzenia 4
Vysvetlivky technických termínov používaných v texte 4
Interaktívna verzia protokolu 4
Spätná väzba 5
Hlavička protokolu 6
Zoznam dokumentov 7
Poznámky k zoznamu 7
Zoznam podobností 8
Plain text 8

Všeobecné údaje protokolu

Systém ANTIPLAG je systém pre podporu rozhodovania. Protokol, ktorý je jeho výstupom:

- nie je potvrdením, že testovaná práca je ORIGINÁL (nemôže odhaliť kopírovanie myšlienky a ani zdrojov ktoré nemá v okamihu porovnania k dispozícii v indexe)
- nie je dôkazom, že testovaná práca je PLAGIÁTOM (napriek vysokému percentuálnemu pokrytia textu testovanej práce môže ísť o text správne citovaný alebo preberaný v súlade so zámerom práce)

Protokol má byť pomôckou. Upozorniť na dokumenty, ktoré mohli uniknúť pozornosti školiteľa alebo oponenta. Cieľom je demonštrovať zhodné a/alebo podobné časti textu kontrolovanej práce voči korpusu prác v systéme.

Strany protokolu sú očíslované. Protokol je chránený heslom proti manipulácii. V päte dokumentu je jednoznačný identifikátor protokolu (PID) vo forme textu. Na prvej strane v záhlaví je PID aj vo forme čiarového kódu (ten je typu Code39). V protokole je vyznačený aj dátum a verzia protokolu a identifikácia portálu v linke umiestnenia vysvetliviek.

Záhlavie protokolu

V záhlaví protokolu je identifikácia testovanej práce. Vychádza z dát dodaných lokálnym úložiskom Záverečných a kvalifikačných prác (ZP) v procese výmeny dát. Zároveň obsahuje údaj o globálnom percente pokrytia textu (označené ako GP). Globálne percento podobnosti nie je jednoduchým súčtom podobností testovaného dokumentu s jednotlivými dokumentami, ktoré boli vyhodnotené ako zaujímavé z pohľadu množstva podobností. Dôvodom je fakt, že časť textov testovanej práce môže byť identifikovaná ako podobná vo viacerých dokumentoch (rovnaký odsek cituje alebo kopíruje viac prác). Celkové percento podobnosti je teda menšie, alebo rovné súčtu parciálnych podobností v jednotlivých dokumentoch (P^1, P^2, ..., P^N) – [$GP <= P^1 + P^2 + ... + P^N$].

Ak zhoda v dokumente presiahne 80%, takýto dokument sa nezapočítava do globálneho percenta (znehodnotil a zakryl by ostatné výsledky), len sa v protokole uvádza jeho bibliografická citácia. Zvyčajne totiž ide o inú verziu dokumentu toho istého autora. Pokrytie textu nad 80% nie je potrebné demonštrovať v detailoch ani prakticky (jedná sa takmer o celý text). Po 1.9.2011 sú špeciálnym prípadom aj práce priamo označené identickým ID zadania (CRZP ID). Zhoda dokumentov s identickým CRZP ID sa rovnako nezapočítava do globálneho percenta a v protokole sa pre poriadok zobrazujú len metadáta a informácia o zhode CRZP ID. Typickým príkladom použitia rovnakého CRZP ID je zasielanie prepracovanej verzie záverečnej práce.*

Grafické zvýraznenie pokrytia textu

Pre zrýchlenie čítania protokolov bolo do protokolu pridané zobrazovanie percentuálnej zhody pomocou farebných štvorčekov. Používatelia systému ANTIPLAG si môžu nastaviť intervaly pre jednotlivé štvorčeky pomocou admin rozhrania (popis nastavenia je súčasťou aktualizovaného manuálu pre admin rozhranie). Intenzita farebnosti zodpovedá zvyšujúcej sa zhode. Veľkosť štvorčeku zodpovedá príslušnosti percenta zhody do nastaveného intervalu.

Čistý text (plain text) - informácie

Od verzie 2.0 protokolu pribudla informácia o niektorých parametroch textu dodanej práce. Čistý text (z angl. plain text) vzniká extrakciou z pôvodného dokumentu a tvorí základ pre ďalšie spracovanie pre

účely porovnávania. Plain text je bez formátovania, značiek, vzorcov alebo obrázkov. Kvalita tohto textu je kľúčová. Pre kontrolu plain textu v protokole pribudlo viacero prvkov:

- informácia o celkovej dĺžke plain textu v znakoch
- informácia o počte slov v plain texte
- informácia o zastúpení slov rôznej dĺžky v plain texte vo vzťahu k typickému priemeru (viď. histogram početnosti slov)

Navyše bol na koniec protokolu pridaný celý plain text práce aby:

- boli nájdené podobnosti zobrazené v celom kontexte práce (vyznačovanie červenou farbou)
- bola možná vizuálna kontrola plain textu pre odhalenie jeho poškodení (úmyselných alebo neúmyselných)

Početnosť slov – histogram

Histogram zobrazuje relatívnu početnosť slov v práci podľa dĺžky slova. Relatívna početnosť je porovnávaná s priemerom a výsledok je označený značkou.

- „=“ – početnosť zodpovedá priemeru
- „>>“ – početnosť je výrazne vyššia ako priemer
- „<<“ – početnosť je výrazne nižšia ako priemer

Priveľa krátkych slov zvyčajne indikuje poškodenie diakritiky alebo problémy kódovania. Priveľa dlhých slov môže indikovať odbornú prácu plnú odborných výrazov, prácu v cudzom jazyku alebo úmyselné poškodzovanie textu vkladaním bielych znakov.

Zoznam dokumentov

Každá práca je porovnávaná voči indexu systému ANTIPLAG. Zo všetkých dokumentov sú vybrané tie, v ktorých sa nachádza nadprahové množstvo podobného textu. Tabuľka „Práce s nadprahovou hodnotou podobnosti“ obsahuje identifikáciu nájdených dokumentov spolu s údajom o percente textového pokrytia. Tento údaj je podielom súčtu dĺžok nájdených odsekov textu a celkovej dĺžky textu testovaného dokumentu (v príklade označené ako P1 a P2). Zoznam je zoradený podľa relevancie. Prvých 5 prác je zobrazených vždy. Z ostatných prác sú zobrazované reprezentatívne vzorky. Sú to práce ktoré do výsledku pridávajú novú informáciu. Teda práca po pozícii N má podobnosť s testovaným dokumentom na pozícií, ktorá nebola identifikovaná žiadnym s dokumentov pred ním (teda na pozícií 1 až N-1).

Podobný text – podobný odsek

Systém je scho+pný hľadať podobnosti už pre sekvencie slov od dĺžky troch plnovýznamových slov bez ohľadu na ich poradie. V praxi sú samozrejme zaujímavé dlhšie sekvencie slov a viet. Systém neoznačuje len texty, ktoré sa vyznačujú absolútnou zhodou. Je schopný riešiť situácie, keď sú slová vo vetách zamenené v poradí, vety sú doplnené o doplnkové slová a podobne.

Minimálna sekvencia je 7 plnovýznamových slov. Štatistická miera podobnosti je nastavená na úroveň 80% podobných plnovýznamových slov. Hovoríme o percente spoľahlivosti odseku a toto číslo je uvádzané napravo od čísla odseku. V prípade zistenia viacerých podobností z rôznych dokumentov na rovnakom mieste dokumentu je spoľahlivosť uvádzaná ako číselný rozsah (príklad: odsek je podobný v dvoch dokumentoch, v jednom so spoľahlivosťou 80%, v druhom 90% a teda spoľahlivosť identifikácie odseku voči týmto dokumentom je v intervale 80 až 90%).

Text citovaný v protokole je text, ktorý sa nachádza v testovanom dokumente (po prevode na plain text). V nájdenom dokumente by sa mal nachádzať text, ktorý je tomuto odseku podobný (slová môžu mať iné poradie, môže ísť o synonymá, slová môžu byť v inom páde, niektoré môžu chýbať, niektoré môžu byť naopak pridané, atď...).

V texte odseku sú vyznačené zarážky (typu »1, «1) určujúce rozsah podobnosti s nájdeným dokumentom. Číslo v zarážke zodpovedá číslu dokumentu v zozname dokumentov. V prípade zjednotenia intervalov z rôznych dokumentov sa čísla dokumentov zobrazujú ako rozsah alebo oddelené čiarkami.

Obmedzenia

- systém je schopný nájsť sekvencie slov, ktoré sú kratšie ako nastavená prahová úroveň, ale neprezentuje ich
- v niektorých prácach sa ako podozrivé vyhodnotia podobné dlhé sekvencie patriace do hlavičky práce alebo do literatúry – tento jav nie je možné úplne potlačiť, ale pozorný čitateľ protokolu vie tento fakt zohľadniť v hodnotení
- slovník plnovýznamových slov bude postupne upravovaný a upresňovaný v súlade s korpusom Slovenského jazyka

Vysvetlivky technických termínov používaných v texte

- **Index** je dátová štruktúra ktorá vzniká spracovaním dodaných dokumentov a vzniká za účelom rýchleho porovnania dokumentov.
- **Plain text** vzniká prevodom originálneho dokumentu na text očistený od značiek, formátovania a ďalších atribútov textu nevhodných pre porovnanie.

Interaktívna verzia protokolu

Od 1.9.2014 systém generuje protokol o kontrole originality verzie 1.5. Hlavné zmeny protokolu:

- Histogram početností je generovaný a prezentovaný aj v nulových protokoloch
- PDF verzia protokolu určená pre tlač obsahuje LINKU na webovú interaktívnu verziu protokolu, linka prakticky kopíruje umiestnenie PDF protokolu vrámci CRZP, len umiestnenie ***/protokol*** nahrádza umiestnením ***/webprotokol***

5FD8F5D[redacted]DDD157

www.crzp.sk/webprotokol?pid=5FD8F5D[redacted]DDD157

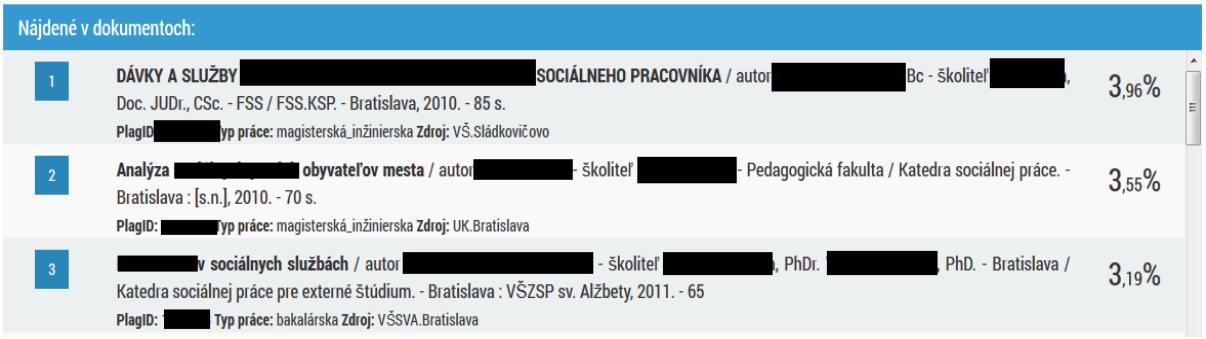
Prvá sekcia protokolu obsahuje anotáciu o používateľovi a kontrolovanom dokumente.

Nájdené podobnosti

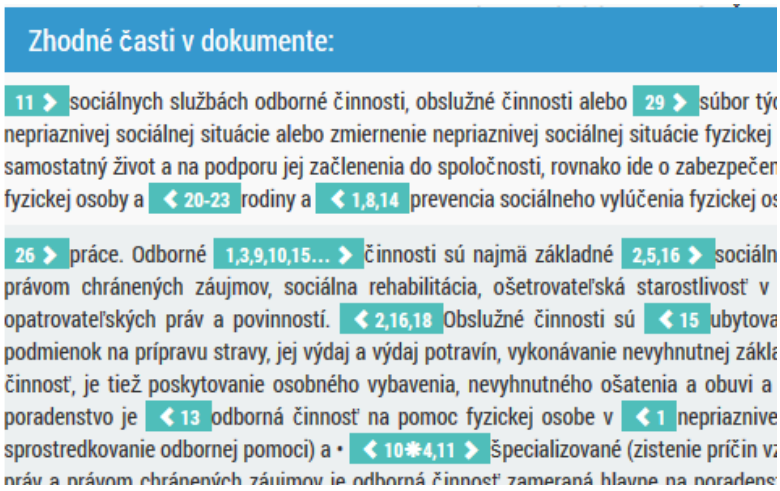
Obec ako poskytovateľ sociálnych služieb [redacted] / autor [redacted], Bc. - školiteľ [redacted] PhDr. Ing., PhD. - FF / [redacted] - Prešov, [redacted] 7,90%

PlagID: [redacted] **Typ práce:** magisterská_inžinierska **Zdroj:** PU.Prešov

V ďaľšej sekcii sa nachádza zoznam odkazov na webové dokumenty, v ktorých bola nájdená zhoda s kontrolovaným dokumentom. Zoradené sú od najvyššej po najnižšiu percentuálnu zhodu. V prvom stĺpci sa nachádza číslo dokumentu. Kliknutím na toto číslo je zvýraznený zhodný text z daného dokumentu v sekcii so zoznamom nájdených zhôd. Kliknutím na odkaz vedľa čísla dokumentu je možné ignorovať všetky nájdené zhody s týmto dokumentom. Anotácia ku zhodným dokumentom obsahuje odkaz na dokument, dátum zaindexovania do systému a jeho veľkosť. V poslednom stĺpci sa nachádza percento zhody s kontrolovaným dokumentom.



Posledná sekcia protokolu obsahuje zoznam nájdených zhôd. Zarážky v texte zhody obsahujú čísla dokumentov, v ktorých bola zhoda nájdené. V prípade, že na určitom intervale začína alebo končí viacero dokumentov, sú čísla dokumentov zobrazené pomocou intervalov (napr. 9-10,15-25). Kliknutím na zarážku je možné ignorovať konrétny zhodný interval.



Spätná väzba

Budeme vďační za pripomienky, návrhy vylepšení, prípadne ohlásenie problémov s protokolom. Tieto informácie prosíme doručovať na adresu grman@svop.sk.

Za tvorivú skupinu ANTIPLAG

Mgr. Ján Grman, PhD.
SVOP spol. s r.o.

22.03.2016, verzia 2.0

Hlavička protokolu

- Identifikátor protokolu vo forme čiarového kódu
- Identifikácia práce vo forme bibliografickej citácie
- Informácie o extrahovanom plain texte
- Histogram početností slov a jeho vzťah k referenčným hodnotám korpusu
    - „=“ v súlade s referenčnou hodnotou
    - „>>“ výrazne viac slov danej šírky
    - „<<“ výrazne menej slov danej šírky

Protokol o kontrole originality

Kontrolovaná práca

Citácia	Percento*
BIBLIOGRAFICKÁ CITÁCIA PRÁCE	**12,56%**

* Číslo vyjadruje percentuálny podiel textu, ktorý má prekryv s indexom prác korpusu CRZP. Intervaly grafického zvýraznenia prekryvu sú nastavené na [0-20, 21-40, 41-60, 61-80, 81-100].

Informácie o extrahovanom texte dodanom na kontrolu

Dĺžka extrahovaného textu v znakoch: 146820
Počet slov textu: 15358

Početnosť slov - histogram

Dĺžka slova	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Indik. odchylka	=	<<	<<	=	=	>>	=	=	=	>>	=	=	=	=	=	=	=	=	=	=	=	=	=

* Odchýlky od priemerných hodnôt početnosti slov. Profil početností slov je počítaný pre korpus slovenských prác. Značka ">>" indikuje výrazne viac slov danej dĺžky ako priemer a značka "<<" výrazne menej slov danej dĺžky ako priemer. Výrazné odchylky môžu indikovať manipuláciu textu. Je potrebné skontrolovať "plaintext" ! Priveľa krátkych slov indikuje vkladanie oddelovačov, alebo znakov netradičného kódovania. Privela dlhých slov indikuje vkladanie bielych znakov, prípadne iný jazyk práce.

Zoznam dokumentov

- Zoznam od najvyššieho percenta
- Prvých 5 vždy (bez ohľadu na to či prinášajú nové intervaly)
- V päte dokumentu je identifikátor, linka na web verziu protokolu, linka na vysvetlivky,verzia a dátum protokolu

Práce s nadprahovou hodnotou podobnosti

Dok.	Citácia	Percento*
1	**Možnosti riešenia konfliktných situácií v činnosti policajno-bezpečnostných orgánov** / autor [redacted] Karol, Bc. - školiteľ [redacted] Ladislav, Ing. - F / KPS. - Bratislava, 2010. - 67 s. [redacted]	2,23%
2	**Postupy a spôsoby konania po eliminácii policajne relevantnej udalosti represívnou metódou verejno-poriadkovej činnosti** / autor Klimčík Peter - školiteľ [redacted] Boris, JUDr. - oponent [redacted] Pavol, Ing., CSc. - F / KPV. - Bratislava, 2013. - 40 s. [redacted]	1,40%
3	**Úloha služby poriadkovej polície v procese eliminácie policajne relevantnej udalosti súvisiace s narušením chráneného objektu** / autor [redacted] Lucia - školiteľ [redacted] Boris, JUDr., PhD. - oponent [redacted] Ladislav, JUDr., PhD. - F / KPV. - Bratislava, 2015. - 73 s. [redacted]	1,34%
4	**Historicko-metodologický vývoj poriadkovej polície** / autor [redacted] Daniel - školiteľ [redacted] Pavol, Ing., CSc. - oponent [redacted] Štefan, JUDr., PhD. - F / KPV. - Bratislava, 2013. - 43 s. [redacted]	1,28%

310[redacted]614201
www.crzp.sk/webprotokol?pid=310[redacted]4201

Poznámky k zoznamu

* Číslo vyjadruje percentuálny prekryv testovaného dokumentu len s dokumentom uvedeným v príslušnom riadku.

Krátené: Dokument má prekryv s veľkým počtom dokumentov. Zoznam dokumentov je krátený a usporiadaný podľa percenta zostupne. Celkový počet dokumentov je **[924]**. V prípade veľkého počtu je často príčinou zhoda v texte, ktorý je predpísaný pre daný typ práce (položky tabuliek, záhlavia, poďakovania). Vo výpise dokumentov sa preferujú dokumenty, ktoré do výsledku prinášajú nový odsek (teda dokumenty ktoré sú plne pokryté podobnosťami iných dokumentov sa v zozname nenachádzajú. Pri prekročení maxima počtu prezentovateľných dokumentov sa v zarážke zobrazuje znak ∞.

Zoznam podobností

2. odsek : spoľahlivosť [78%]

[18»] Odraz sociálnych konfliktov v činnosti policajno-bezpečnostných orgánov. In Policajná teória a prax, 1996. roč. 4. č. 3. 22 s. 2 HAŠEK, J. 2005. Interpersonálne vzťahy v Policajnom zbore a psychologické aspekty ich upevňovania. In Policajná teória a prax, 2005. roč. 13. č. 3, 29 s. 3[«18] ČAKRT,

3. odsek : spoľahlivosť [89%]

[401»] charakter. Psychológia uvádza, že konfliktná situácia je zvláštnym druhom medziosobnostného styku medzi dvoma, prípadne viacerými osobnosťami, pri ktorom dochádza k stretu ich postojov, záujmov, cieľov, predstáv, alebo hodnôt. Do správania a jednania účastníkov konfliktu sa vtedy premieta vzniknutý interpersonálny konflikt a je základom konfliktnej situácie.5[«401] Konflikty

4. odsek : spoľahlivosť [92%]

[11»]„fiktívny“ subjekt rozhodovania. Naproti tomu, psychologický prístup vždy vychádza zo skutočnosti, že subjekt rozhodovania je konkrétny človek, či skupina ľudí so svojimi názormi, schopnosťami, cieľmi a záujmami; so svojimi prednosťami, či nedostatkami. A práve tieto nedostatky sú označované ako „obmedzená racionalita subjektu rozhodovania“ a jej faktory môžeme rozdeliť do[«11] nasledujúcich

5. odsek : spoľahlivosť [86%]

[396»] zákrok, prípadne iný úkon, prejavuje z hľadiska podmienok riešenia, východísk, ale aj vo vzťahu k možným dôsledkom tohto riešenia – a to bezprostredným alebo sprostredkovaným.) 3. Nekonzistentnosť preferencií u rozhodovateľa – prejavujúca sa v oblasti preferencie cieľov, hodnôt a variant riešenia konkrétneho problému. 4. Vnútorný konflikt rozhodovací – tento sa prejavuje v[«396] tom,

6. odsek : spoľahlivosť [84%]

[93»] ľudstvo sa delí na vyššie rasy a nižšie rasy, pričom vyšším rasám sa vraj priznáva tvorivosť ako aj to, že sú nositeľmi pokroku i civilizácie a určené sú na vládnutie. Naopak nižšie rasy je potrebné viesť a[«93] chýba

Plain text

Plain text dokumentu na kontrolu

Skontroluje extrahovaný text práce na konci protokolu! Plain text (čistý text - extrahovaný text) dokumentu je základom pre textový analyzátor. Tento text môže byť poškodený úmyselne (vkladaním znakov, používaním neštandardných znakových sád, ...) alebo neúmyselne (napr. pri konverzii na PDF nekvalitným programom). Nepoškodený text je čitateľný, slová sú správne oddelené, diakritické znaky sú správne, množstvo textu je primeraný rozsahu práce. Pri podozrení na poškodený text (väčšieho rozsahu), je potrebné prácu na kontrolu originality zaslať opakovane pod rovnakým CRZPID.

Spôsoby možného riešenia konfliktných situácií orgánmi polície

V našej spoločnosti stále pretrváva veľké množstvo negatívnych javov, .ktoré súvisia s absenciou morálnych noriem v správaní a konaní u niektorých z nás. Čoraz častejšie sa stretávame s narušovaním verejného poriadku, so zvyšujúcou sa nedisciplinovanosťou niektorých jedincov i skupín, ktorí pohŕdajú štátnymi orgánmi a nerešpektujú zákon. Spáchané trestné činy ako aj ich páchatelia bývajú viac utajení, prípadne kvalifikovanejší. Hrubosť,[25»]neúcta, bezohľadnosť a celková negatívna hodnotová orientácia iba na zisk a majetok „za každú cenu“ často sprevádzajú a poznačia medziľudské vzťahy. U občanov sa znižuje dôvera k policajnému zboru, klesá u nich pocit bezpečnosti. Polícia je inštitúciou, ktorá predstavuje nástroj verejnej moci. Chráni práva a oprávnené záujmy fyzických a právnických osôb, ústavné zriadenie, život, zdravie a majetok každého z nás. Taktiež zabezpečuje vnútornú funkciu štátu, ktorá je zameraná na ochranu bezpečnosti a verejného poriadku. Zákon jej stanovuje právomoci, ktoré pri výkone svojej činnosti využíva.19[«25]Policajt v[19»]spoločnosti vystupuje ako verejný[25»]činiteľ a súčasne orgán činný v trestnom konaní. Z hľadiska výkonu svojej služby a zamerania sa pomerne často dostáva do konfliktných a nebezpečných situácií. Povaha jeho práce si vyžaduje kontakt s relatívne obmedzenou, špecifickou časťou populácie, ktorá otvorene prejavuje svoj negatívny postoj k polícii a políciu odmieta. Konfliktné a nebezpečné situácie vyvolávajú najčastejšie osoby, ktoré sa neidentifikujú so stanovenými pravidlami a normami,[«19]sústavne ich porušujú a vôbec ich nerešpektujú. Policajti sa tiež kontaktujú s osobami, ktoré klamú, zavádzajú ich, nehovoria pravdu a týmto sa dostávajú do[«25]antipatických sociálnych interakcií s inými jednotlivcami.[67»]Bežné konflikty patria k existencii človeka v jeho reálnom svete; sú mimo neho, aj v ňom samom. Sú produktom rozporov medzi spontánnymi tendenciami – „to, čo chce“ a sociálnymi podmienkami ich realizácie – „to, čo môže“.[«67]Všeobecne sa pre riešenie konfliktných situácií odporúča postupovať podľa tzv. „zdravého rozumu“. Nie každú konfliktnú situáciu môžeme popísať úplne presne a stanoviť presný a konkrétny postup, alebo návod, ako ju riešiť. Každá je individuálna, odlišná od inej a policajt ich musí riešiť v závislosti od okolností rozdielnymi spôsobmi. Preto je významné, poznať nové orientácie v chápaní konfliktov, ktoré sa v čistej podobe prejavili vo vzniku

19

VLKOVÁ, V. 2002. Psychologické hľadiská úspešnej komunikácie policajta v konfliktných situáciách. Bratislava: odbor vzdelávania a psychológie SPSČ MV SR, 2002. 2 s.

27