

# Ako čítať a interpretovať „Protokol o kontrole originality“

(pre protokoly systému CRZP/ANTIPLAG verzia 3.0, aktualizácia: 23.8.2022)

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. Okrem protokolu určeného pre tlač (v súbore PDF), existuje aj interaktívna verzia protokolu. Táto je dostupná z linky ktorá je umiestnená na konci protokolu.

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

## Obsah

Vysvetlivky k obsahu protokolu.....	2
Všeobecné údaje protokolu .....	2
Ochranné prvky protokolu .....	2
Záhlavie protokolu.....	3
Grafické zvýraznenie pokrytia textu.....	3
Slovníková analýza textu .....	4
Čistý text (plain text) – frekvenčná analýza.....	5
Početnosť slov – histogram .....	5
Zoznam dokumentov.....	5
Podobný text – podobný odsek.....	6
Obmedzenia .....	6
Interaktívna verzia protokolu .....	6
Offline kópia analýzy protokolu .....	8
Štruktúra protokolu .....	8
Záhlavie protokolu.....	8
Zoznam dokumentov.....	8
Poznámky k zoznamu .....	9
Päta protokolu.....	9
Zoznam podobností.....	10
Plain text.....	10
Vysvetlivky technických termínov používaných v texte .....	11

## Vysvetlivky k obsahu protokolu

### Všeobecné údaje protokolu

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

- nie je potvrdením, že testovaná práca je **ORIGINÁL**
- nie je dôkazom, že testovaná práca je **PLAGIÁTOM**

Protokol má byť pomôckou. Upozorniť na dokumenty, ktoré mohli uniknúť pozornosti školiteľa a oponenta. Navyiac korpus nemôže obsahovať všetky dokumenty z ktorých autori mohli čerpať. Mnoho zdrojov nemá elektronickú formu, alebo nie sú dostupné na internete. Cieľom je demonštrovať zhodné a/alebo podobné časti textu kontrolovanej práce voči korpusu prác v systéme.

### Nízke percento zhody

Protokol s nízkym percentom nevylučuje plagiátorstvo. Prečo?

- Formálna kontrola nemôže odhaliť kopírovanie myšlienky. Ak je myšlienka formulovaná inými slovami, antiplagiátorské algoritmy nedokážu odhaliť zhodu myšlienky ale len zhodu alebo podobnosť textu.
- Algoritmus pracuje na základe porovnávania nového textu s obsahom získaných a indexovaných zdrojov. Existuje mnoho textov ktoré nie sú verejné. Ak študentovi prácu napíše iný človek a jeho práca je originálna, nie je možné to formálne zistiť.
- Špeciálne je potrebné sa zamerať na protokoly s nulovou alebo veľmi nízkou zhodou. Je potrebné pozorne prezrieť text práce uvedený na konci protokolu, či sa zhoduje s tlačným originálom. Školiteľ vie tiež odhaliť nadmerné parafrázovanie zdrojov.

### Vysoké percento zhody

Vyššie percento v protokole nie je dôkazom plagiátorstva. Zhoda textu môže byť zapríčinená:

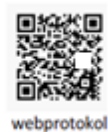
- Viaceré práce rovnakého autora boli zaslané do CRZP opakovane – niekedy omylom
- Študent používa presný text ako formu argumentácie. Niektoré typy textov (zákony, normy) nie je vhodné parafrázovať.
- Študent nadmerne cituje. Správne cituje, ale zhodného textu je veľa. Pri veľkom objeme zhodného textu môže byť oprávnená obava o rozsah vlastného vkladu študenta.

### Ochranné prvky protokolu

Strany protokolu sú očíslované. Protokol je chránený heslom proti manipulácii. V hlavičke dokumentu sú dva QR kódy.



## Protokol o kontrole originality



Obrázok: identifikátory

OR kód vľavo reprezentuje URL linku na verejný portál CRZP na detail záznamu pre ktorý bol protokol vydaný. Ide o jednoduchú kontrolu faktu, že údaje protokolu korešpondujú s údajmi v databáze CRZP.




QR kód vpravo reprezentuje URL linku na verejný portál CRZP na funkciu pre analýzu protokolu. V tejto časti portálu je možné dynamicky vypínať a zapínať jednotlivé časti protokolu a spúšťať alternatívny výpočet percenta zhody.

## Záhlavie protokolu

V záhlaví protokolu je identifikácia testovanej práce. Vychádza z dát dodaných VŠ a jej 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 ( $P^1, P^2, \dots, P^N$ ) –  $[GP \leq P^1 + P^2 + \dots + P^N]$ .

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.\*

## Kontrolovaná práca

Citácia	Percento*
 https://opac.crzp.sk, 2022. - PDF 1639 kb plagID:  zdroj: CVTI.Bratislava	<b>4,68%</b> 

\* 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].

**Zhoda v korpusoch:** Korpus CRZP: 0,00% (0), Internet: 0,87% (63), Wiki: 4,67% (70), Slov-Lex: 0,00% (0)

Obrázok: záhlavie (kontrolovaná práca) + zhody (globálne percento a zhoda v korpusoch)

**Novinka:** V protokole od verzie 3.0 je okrem globálneho percenta zhody vyčíslená zhoda aj v jednotlivých význačných korpusoch dokumentov (text zhoda v korpusoch). Pre tieto percentá platí znovu množinová matematika. Nejde o jednoduchý súčet. Ak študent citoval text z Wikipédie a rovnaký text citovali aj iní študenti, nájdený zhodný text sa prejaví aj v korpuse Wiki aj v korpuse CRZP. Za percentom v zátvorke je počet dokumentov so zhodou.

## 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 štvorčky 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.

## Slovníková analýza textu

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.

V systéme ANTIPLAG bol vybudovaný referenčný slovník slov. Tento slovník obsahuje kontrolovaný súbor slov v jazykoch SK, CZ, EN, DE, HU. Vďaka jeho existencii bolo možné vytvoriť nový typ testu kontrolovanej práce pred manipuláciou. Text práce je tak možné okrem frekvenčnej analýze dĺžky slov možné porovnávať aj s týmto referenčným slovníkom. Výsledky tejto analýzy sú v protokole od verzie 3.0.

Čím je zhoda slov s referenčným slovníkom nižšia, tým viac platí

- Spoľahlivosť porovnávania voči korpusu je nižšia
- Porovnávaný dokument obsahuje text ktorý je (platí jedna alebo viac podmienok)
  - manipulovaný (vkladaním znakov, neviditeľných textov a podobne)
  - v nepodporovanom jazyku
  - nesprávne prevedený do PDF a neprevoditeľný na správny čistý text (viditeľný text práce v PDF nezodpovedá strojovo extrahovanému)

## Informácie o extrahovanom texte dodanom na kontrolu

Dĺžka extrahovaného textu v znakoch: 102882

Počet slov v slovníku (SK, CZ, EN, HU, DE): 9238

Súčet dĺžky slov v slovníku (SK, CZ, EN, HU, DE): 70646

Celkový počet slov textu: 13736

Pomer počtu slovníkových slov: 67,3%

Pomer dĺžky slovníkových slov: 68,7%

Interval	100%-70%	70%-60%	60%-50%	40%-30%	30%-0%
Vplyv na KO*	žiadny	malý	stredný	veľký	zásadný

\* Kontrola originality je výrazne oplyvnená kvalitou dodaného textu. Slovníkový test vyjadruje mieru zhody slov kontrolovanej práce so slovníkom referenčných slov podporovaných jazykov. Nízka zhoda môže byť spôsobená: nepodporovaný jazyk, chyba prevodu PDF alebo úmyselná manipulácia textu. Text práce na vizuálnu kontrolu je na konci protokolu.

Obrázok: slovníková analýza textu

Pri indikácii vplyvu na **kontrolu originality (KO)**

- **v stupni stredný** odporúčame prezrieť text na konci protokolu a vizuálne ho porovnať s textom v PDF, vplyv na výsledky kontroly a skreslenie je do 10%
- **v stupni veľký a zásadný** odporúčame prezrieť text pre určenie dôvodov (manipulácia alebo jednoducho dodaný zlý formát PDF), rozhodne však výsledky kontroly nie je možné považovať za relevantné a ak je možné problém na strane študenta odstrániť, odporúčame zaslať opravenú prácu znovu na kontrolu

Pred každým zaslaním sa odporúča študentom používať službu <https://testdoc.crzp.sk> na kontrolu textu PDF. Kontrola obsahuje aj kontrolu voči referenčnému slovníku.

## Čistý text (plain text) – frekvenčná analýza

Od verzie 2.0 protokolu pribudla informácia o niektorých parametroch textu dodanej práce. Ide o tieto prvky:

- 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)

Navýše 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 vkladáním bielych znakov.

### 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 početnosti 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 odchýlky môžu indikovať manipuláciu textu. Je potrebné skontrolovať "plaintext"! Priveľa krátkych slov indikuje vkladanie oddeľovačov, alebo znakov netradičného kódovania. Priveľa dlhých slov indikuje vkladanie bielych znakov, prípadne iný jazyk práce.

Obrázok: frekvenčná analýza textu

## 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 pozícii, ktorá nebola identifikovaná žiadnym s dokumentov pred ním (teda na pozícii 1 až N-1).

## Podobný text – podobný odsek

Systém je schopný 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 je upravovaný a upresňovaný v súlade s korpusom Slovenského jazyka

## Interaktívna verzia protokolu

Spolu s verziou 3.0 papierového protokolu prichádza v roku 2022 aj úplne nový interaktívny protokol. V tomto obsahovo identickom protokole je snaha o maximálnu podobnosť s papierovým a zároveň možnosť vyradovať z porovnávania celé dokumenty alebo nájdené sekvencie textu. Táto činnosť umožňuje používateľovi napríklad vyradiť z výpočtu percenta zhody zjavne správne citované dokumenty alebo ich časti. Tým vo výsledku zostanú len nesprávne citované texty, alebo texty považované za možné plagiátorstvo. Prepočítané percento tak potom dáva reálnejší obraz o miere nesprávneho citovania alebo plagiátorstva.

Prostredie práce s protokolom je úplne nové a prepracované. V hornej časti je citácia kontrolovanej práce. V ľavej časti je stránkovaný zoznam nájdených dokumentov s podobnosťami textu. V pravej časti je zoznam častí textu – táto časť sa naplní po zvolení konkrétneho dokumentu vľavo. Aj zoznam častí textov je stránkovaný.

**Holistický koncept budovania a riadenia značky v priemyselnom podniku v podmienkach Marketingu 4.0** / Autor: Babčanová, Dagmar -  
 Oponent: Chromjaková, Felicitia - Oponent: Dušina, Ľuboslav - Oponent: Závadský, Ján  
 PlagID: 1594237 Typ práce: habilitačná Zdroj: Slovenská technická univ. v Bratislave MTF UPIM (MTF)

13,88 %

## Nájdene v dokumentoch

<input checked="" type="checkbox"/>	<b>Návrh metodiky budovania a riadenia značky v priemyselných podnikoch</b> / Autor: Babčanová, Dagmar - Školiteľ: Čambál, Miloš PlagID: 1056662 Typ práce: dizertačná Zdroj: Slovenská technická univ. v Bratislave MTF UPIMK (MTF)	1,46 %
<input checked="" type="checkbox"/>	<a href="http://of.euba.sk/merkur_2012_cd/prispevky/ZZborn%C3%ADk%20KONIEC.pdf">http://of.euba.sk/merkur_2012_cd/prispevky/ZZborn%C3%ADk%20KONIEC.pdf</a>	1,35 %
<input checked="" type="checkbox"/>	<a href="http://of.euba.sk/merkur_2012_cd/prispevky/%C5%A0il%C3%A1rov%C3%A1_M...">http://of.euba.sk/merkur_2012_cd/prispevky/%C5%A0il%C3%A1rov%C3%A1_M...</a>	1,35 %
<input checked="" type="checkbox"/>	<a href="http://merkur.obchodnafakulta.sk/zborniky/zbornik2012.pdf">http://merkur.obchodnafakulta.sk/zborniky/zbornik2012.pdf</a>	1,35 %
<input checked="" type="checkbox"/>	<b>Moderné prístupy a metódy v marketingovom manažmente</b> / Autor: Šilhárová, Marta - Školiteľ: Daňo, Ferdinand - Oponent: Bugár, Štefan - Oponent: Kita, Jaroslav - Oponent: Pajtíková Bartáková, Gabriela PlagID: 1362009 Typ práce: dizertačná Zdroj: Ekonomická univerzita v Bratislave OF KMr OF	1,22 %

1 / 288

## Nájdene časti dokumentu

Časti textu záznamu/linku	
<input checked="" type="checkbox"/>	potrebné expandovať. Rozmáhajúca sa globalizácia ovplyvňuje marketingové aktivity, produkty tiež nároky a potreby priemyselných podnikov. Pre budúci úspech je nevyhnutné vytvoriť a následne vhodne riadiť silné značky, ktoré vplyvajú
<input checked="" type="checkbox"/>	hlinené lampy sa často predávali priamo v obchodoch, kde boli aj vyrábané, a kupujúci vyhľadávali značku spoľahlivých hlinčiarov, ktorí boli pre nich zárukou kvality. Značky možno nájsť už v ranom čínskom porceláne, na hlinených džbánoch z prastareho Grécka a Rima alebo na tovaroch ...vac
<input checked="" type="checkbox"/>	vznikom balených výrobkov. Industrializácia presunula produkciu veľkého množstva výrobkov pre domácnosti z menších lokálnych podnikov do centralizovaných tovární. V polovici
<input checked="" type="checkbox"/>	značku nasledovne: „Značka je ľudskými zmyslami ťvminateľné označenie organizácie a jej produktov, prostredníctvom ktorého je zákazník schopný rozlíšiť konkrétnu organizáciu a jej produkty od iných. Značka organizáciu a jej produkty identifikuje a zároveň diferencuje ...vac
<input checked="" type="checkbox"/>	známky Označenie, ktoré môže tvoriť ochrannú známku, je akékoľvek označenie, ktoré možno graficky znázorniť a ktoré tvoria najmä slová vrátane osobných mien, písmená, číslice, kresby, tvar tovaru alebo jeho obal, prípadne ich vzájomné kombinácie, ak takéto označenie je spôsobilé ...vac

Obrázok: základné prostredie interaktívneho protokolu

Je možné vyradiť z počítania percent buď celý dokument vyznačením VĽAVO

Nájdene v dokumentoch		
<input type="checkbox"/>	<b>Návrh metodiky budovania a riadenia značky v priemyselných podnikoch</b> / Autor: Babčanová, Dagmar - Školiteľ: Čambál, Miloš PlagID: 1056662 Typ práce: dizertačná Zdroj: Slovenská technická univ. v Bratislave MTF UPIMK (MTF)	1,46 %
<input checked="" type="checkbox"/>	<a href="http://of.euba.sk/merkur_2012_cd/prispevky/ZZborn%C3%ADk%20KONIEC.pdf">http://of.euba.sk/merkur_2012_cd/prispevky/ZZborn%C3%ADk%20KONIEC.pdf</a>	1,35 %

Obrázok: označenie vyradeného protokolu z výpočtu percenta zhody

Vyradiť je možné len istý nájdený odsek vyznačením odseku VPRAVO.

<input checked="" type="checkbox"/>	<b>Moderné prístupy a metódy v marketingovom manažmente</b> / Autor: Šilhárová, Marta - Školiteľ: Daňo, Ferdinand - Oponent: Bugár, Štefan - Oponent: Kita, Jaroslav - Oponent: Pajtíková Bartáková, Gabriela PlagID: 1362009 Typ práce: dizertačná Zdroj: Ekonomická univerzita v Bratislave OF KMr OF	1,22 %
<input checked="" type="checkbox"/>	marketingVnútroty marketing smeruje k zamestnancom organizácie na všetkých úrovniach a je zameraný na obojstrannú komunikáciu so zákazníkmi a ostatnými účastníkmi dodávateľského reťazca. Na dosiahnutie čo najvyššieho stupňa efektivity je nevyhnutný nábor schopných ...vac	
<input type="checkbox"/>	marketing sa stará o to, aby sa pri marketingových rozhodnutiach brali do úvahy finančné záležitosti, hodnota pre zákazníka, samotná značka, spoločnosť, zákony, etika, rôzne komunity, bezpečnosť a v neposlednom rade dopady na životné prostredie. Marketingovi manažer ...vac	

Obrázok: označenie vyradeného odseku textu z jedného z dokumentov

Protokol sa po každej zmene prepočíta, pričom bol kompletne zmenený spôsob vyradovania textu:

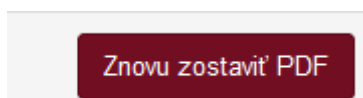
- Vyradením dokumentu je vyradený celý zoznam v ňom nájdených textových intervalov
- Vyradený odsek je jeden textový interval.
- Textový interval je v skratke definovaný ako výsek textu OD – DO (ide o absolútnu pozíciu v texte kontrolovanej práce)
- Pôvodný algoritmus vyradil tieto intervaly a pokrytie textu počítal z ostatných intervalov. Predpokladajme, že sme vyradili text od pozície 100 do pozície 1500. Ak v inom dokumente

bola nájdená zhoda od pozície 1 do 3000, tak to spôsobilo, že vyradením intervalu 100 až 1500 sa percento neznižilo. Text bol naďalej pokrytý tým dlhším a nevyradeným intervalom.

- Nový algoritmus od verzie 3.0 vyradený interval aplikuje na zjednotený text. Vo vyššie uvedenom príklade dvoch nájdených intervalov 100-1500 a 1-3000 je výsledok pri vyradení 100-1500 počítaný ako zjednotenie intervalov 1-99 a 1501-3000.
- Tento spôsob výpočtu viac zodpovedá očakávaniam používateľa. Je nám jedno v koľkých opakovaniach bol rovnaký text identifikovaný, my ho označujeme ako správne citovaný a chceme ním ponížiť percento.

## Offline kópia analýzy protokolu

Novinkou je implementácia funkcie pre uloženie analýzy protokolu do PDF formátu. V pravom hornom rohu okna je tlačidlo



Táto funkcia vytvorí protokol podobného dizajnu ako má oficiálne generovaný protokol. Odlišné je záhlavie protokolu. V protokole sú vyznačené odstránené dokumenty a intervaly (prečiarknutý text). V plaintexte sa vyradený text neukazuje farebne. Percentá sú samozrejme zmenené. Takýto protokol môže slúžiť ako argumentačná báza pre školiteľa alebo študenta v rámci obhajoby záverečnej práce. Protokol je šifrovaný a nie je možné jeho elektronickú verziu upraviť a zneužiť. Pri tlačných verziách je potrebné dávať pozor a preferovať elektronické zaslanie.



metadata

## Analýza protokolu o kontrole originality

### Kontrolovaná práca

Obrázok: ukážka záhlavia analytickej verzie protokolu

## Štruktúra protokolu

### Záhlavie protokolu

- Identifikátor protokolu vo forme QR kódov s URL linkami na metadáta a interaktívny protokol
- Identifikácia práce vo forme bibliografickej citácie
- Zhody v korpusoch
- Slovníková analýza textu
- Frekvenčná analýza textu




### Zoznam dokumentov

- Zoznam od najvyššieho percenta
- Prvých 5 dokumentov je citovaných vždy, bez ohľadu na to či prinášajú nové zhody



- Zvyšok zoznamu práce také dokumenty neobsahuje. Ak napríklad dokument na 12tej pozícii obsahuje len zhody už prezentované v dokumentoch 1 až 11, potom sa tento dokument v zozname nenachádza. Sú práce ktoré majú zhody v tisícoch dokumentov.

### Práce s nadprahovou hodnotou podobnosti

Dok.	Citácia	Percento*
1	<a href="https://fses.uniba.sk/fileadmin/fsev/o_fakulte/dokumenty/kniznica_2015_16/metodicke-pokyny-crepc2_2018_0903.pdf">https://fses.uniba.sk/fileadmin/fsev/o_fakulte/dokumenty/kniznica_2015_16/metodicke-pokyny-crepc2_2018_0903.pdf</a> / Stiahnuté: 19.12.2018; Veľkosť: 237,66kB. <i>plagID: 46318829 zdroj: internet/intranet</i>	1,94% 
2	<a href="https://www.upjs.sk/public/media/2731/metodicke-pokyny-crepc2_2018_0903.pdf">https://www.upjs.sk/public/media/2731/metodicke-pokyny-crepc2_2018_0903.pdf</a> / Stiahnuté: 12.07.2019; Veľkosť: 237,66kB. <i>plagID: 49237665 zdroj: internet/intranet</i>	1,94% 
3	<a href="http://cms.crep.sk/Data/Sites/1/pdf/metodickepokyny/metodicke-pokyny-crepc2-2018-0903.pdf">http://cms.crep.sk/Data/Sites/1/pdf/metodickepokyny/metodicke-pokyny-crepc2-2018-0903.pdf</a> / Stiahnuté: 01.08.2019; Veľkosť: 237,74kB. <i>plagID: 49384956 zdroj: internet/intranet</i>	1,94% 

### 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 ∞.

### Päta protokolu

- V päte strán je číslo strany linka na vysvetlivky,verzia a dátum protokolu
- Na konci protokolu je linka na metadáta a interaktívny protokol v textovej forme (rovnaké URL ako kódované v QR kódach)

## Zoznam podobností

<b>2. odsek :</b>	<b>spoľahlivosť [78%]</b>
<b>[18»]</b> Odras 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,	
<b>3. odsek :</b>	<b>spoľahlivosť [89%]</b>
<b>[401»]</b> charakter. Psychológia uvádza, že konfliktná situácia je zvláštnym druhom medziosobného 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	
<b>4. odsek :</b>	<b>spoľahlivosť [92%]</b>
<b>[11»]</b> „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	
<b>5. odsek :</b>	<b>spoľahlivosť [86%]</b>
<b>[396»]</b> zárok, 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. Vnútorňý konflikt rozhodovací – tento sa prejavuje v[«396] tom,	
<b>6. odsek :</b>	<b>spoľahlivosť [84%]</b>
<b>[93»]</b> 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 pretrvávajú 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 pohrdajú 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 medzilidské 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]Policajti 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 polícii odmieta. Konfliktne 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 policajti 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 SPŠ MV SR, 2002. 2 s.

27

## Vysvetlivky technických termínov používaných v texte

- **ANTIPLAG** je konkrétna implementácia systému pre odhaľovania plagiátorstva od spoločnosti SVOP spol. s r.o. pre účely Centrálného registra záverečných a kvalifikačných prác v Slovenskej republike (CRZP).
- **Korpus** je archív dokumentov z rôznych zdrojov a formátov, z ktorých je možné získať ich elektronickú textovú podobu a slúži ako podklad pre porovnávanie novo vloženého dokumentu.
- **Index** je sofistikovaná 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.
- **Histogram** v kontexte protokolu o kontrole originality je relatívna (alebo aj normalizovaná) početnosť slov dĺžky 3 až 25 v dodanom texte. Získané histogramy sa porovnávajú s referenčným histogramom získaným analýzou spoľahlivého výberu referenčných dokumentov (spoľahlivého z pohľadu jazykovo správneho textu).
- **Slovníkové slovo** je slovo, ktoré sa nachádza v referenčnom slovníku
- **Pomer počtu slovníkových slov** je percentuálna miera (pomer) počtu slovníkových a počtu všetkých slov v dokumente.
- **Pomer dĺžky slovníkových slov** je percentuálna miera (pomer) súčtu dĺžok slovníkových a súčtu dĺžok všetkých slov v dokumente
- **Nadprahové množstvo textu** je termín ktorým označujeme systém hodnotiacich kritérií ktoré sa nastavujú v jadre antiplagiátorského algoritmu na rôznych úrovniach jeho funkcionality za účelom rozhodnutia, či dve slova, dva odseky, dva dokumenty sú zhodné alebo podozrivo podobné. Rozhodnutia sú vždy založené na technických parametroch, ktorých hodnoty boli získané štatistickou analýzou, experimentom a validáciou dokumentov v korpuse systému