

Ako čítať a interpretovať „Protokol o kontrole originality“

(pre protokoly systému CRZP)

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

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

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

Protokol má byť pomôckou. Upozorňuje na dokumenty, ktoré mohli uniknúť pozornosti školiteľa alebo oponenta.

Všeobecné údaje protokolu

Strany protokolu sú očíslované. Protokol je chránený heslom proti manipulácii. V päte dokumentu je jednoznačný identifikátor protokolu (PID) vo forme textu aj čiarového kódu (ten je typu Code39), identifikácia portálu v linke umiestnenia na vysvetlivky, dátum vytvorenia a verzia protokolu.

Záhlavie protokolu

V záhlaví protokolu je identifikácia kontrolovanej práce. Vychádza z dát dodaných lokálnym úložiskom Záverečných a kvalifikačných prác (ZP) v procese výmeny dát. Zároveň obsahuje údaj o globálnom percente prekryvu textu (označené ako GP). Globálne percento podobnosti nie je jednoduchým súčtom podobností kontrolovanej práce s jednotlivými dokumentmi, ktoré boli vyhodnotené ako zaujímavé z pohľadu množstva podobností. Dôvodom je fakt, že časť textov kontrolovanej 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 70%, 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. Prekryv nad 70% 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.

Grafické zvýraznenie prekryvu

Pre zrýchlenie čítania protokolov bolo do protokolu pridané zobrazovanie percentuálnej zhody pomocou farebných štvorčekov. Používatelia systému ANTIPLAG (vysoké školy) 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 štvorčeka zodpovedá príslušnosti percenta zhody nastaveného intervalu.

Početnosť slov – histogram

Histogram zobrazuje relatívnu početnosť slov v práci podľa dĺžky slova. V prípade, že práca obsahuje neštandardnú početnosť slov určitej dĺžky, tak je táto početnosť zvýraznená inou farbou. Môže to znamenať napr. to, že „plain“ text práce nie je možné správne prekonvertovať a obsahuje zhluky nečitateľných znakov alebo slov.

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 prekryvu. Tento údaj je podielom súčtu dĺžok nájdených odsekov textu a celkovej dĺžky textu kontrolovanej práce (v príklade označené ako P1 a P2).

Pokiaľ má kontrolovaná práca prekryv s N dokumentmi s nadprahovým množstvom podobného textu a pritom $N > 15$, potom sa v protokole vytlačí len prvých 15 dokumentov s najvyšším prekryvom. Do percenta zhody sa zarátavajú prekryvy všetkých nadprahových dokumentov.

Podobný text – podobný odsek

System 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. V tomto smere je systém parametrizovateľný a tieto parametre sa môžu upresňovať pre rôzne typy dokumentov. 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.

Detaily – zistené podobnosti

Z praktických dôvodov (aby protokoly nemali veľký rozsah), ale aj z ideových dôvodov (je potrebné identifikovať a prezentovať ako problém jedinú izolovanú vetu?) sa definuje minimálny počet plnovýznamových slov, ktoré sa musia nachádzať v odseku označenom ako podozrivý. Navyše tieto slová musia na „ploche“ tohto odseku tvoriť nadprahové percento textu odseku. Hovoríme o percente spoľahlivosti odseku a toto číslo je uvádzané napravo od čísla odseku. V prípade prekryvu zistených podobností v odseku je spoľahlivosť uvádzaná ako číselný rozsah (odsek je podobný v dvoch dokumentoch, v jednom so spoľahlivosťou 80%, v druhom 90% a teda spoľahlivosť identifikácie odseku voči týmto dokumentom je 80 až 90%).

Text citovaný v protokole je textom, ktorý sa nachádza v kontrolovanej práci (po prevode na plain text). V nájdenom dokumente by sa mal nachádzať text ktorý je tomuto odseku podobný (slová môžu mať iné poradie, 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 prekryvu sa čísla dokumentov zobrazujú ako rozsah (»1-5), oddelené čiarkami (»1,3,4) alebo kombinovane (»1,3,5-8). Znak » určuje začiatok a znak « koniec rozsahu. Napr. zarážka [»1,3,5-8] určuje začiatok rozsahu podobnosti s nájdenými dokumentmi č. 1, 3 a 5 až 8.

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 by bolo možné redukovať odsekávaním začiatkov a koncov dokumentov – v tejto fáze projektu sme k tomuto kroku zatiaľ nepristúpili,
- 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.

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.

1.3.2013, verzia 1.4

Protokol o kontrole originality

Kontrolovaná práca

Citácia	Percento*
BIBLIOGRAFICKÁ IDENTIFIKÁCIA TESTOVANÉHO DOKUMENTU	7,30% □□□□□

← GP

* Číslo vyjadruje percentuálny podiel textu, ktorý má prekryv s indexom prác korpusu CRZP. Intervaly grafického zvýraznenia prekryvu sú nastavené na [0-11, 12-37, 38-62, 63-84, 85-100].

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
Relat. početnosť	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

* Neštandardné odchýlky od priemerných hodnôt sú vyznačené červenou farbou.

Práce s nadprahovou hodnotou podobnosti

Dok.	Citácia	Percento*
1	BIBLIOGRAFICKÁ IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU Č.1	85,04%^{*2} □□□□■ ← P1
2	BIBLIOGRAFICKÁ IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU Č.2	5,89%^{*3} □□□□□ ← P2
3	BIBLIOGRAFICKÁ IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU Č.3	2,87% □□□□□ ← P3
4	BIBLIOGRAFICKÁ IDENTIFIKÁCIA NÁJDENÉHO DOKUMENTU Č.4	2,59% □□□□□ ← P4

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

*2 Nadlimitný prekryv dokumentov: Dokument vykazuje vysokú percentuálnu zhodu textu. Často sa jedná o dvakrát odovzdanú prácu rovnakého autora. Pre zamedzenie skreslenia výsledku tento dokument nie je zarátaný do celkového výstupu.

*3 Dokument má zhodný identifikátor s porovnávanou prácou a preto nebol zarátaný do celkového výstupu.

Identifikátor protokolu a dokumentu v CRZP
(aj vo forme čiarového kódu)

Dátum vygenerovania protokolu a verzia



10B09D204743411D86B740B2A82C3691

20.08.2012 (verzia 1.3)

- 1 -

www.crzp.sk/vysvetlivky.pdf

Detaily - zistené podobnosti

Percento spoľahlivosti (PS)



1. odsek :	spoľahlivosť 100%
[2 »] podstaty trestného činu 3.1 Obligatórne znaky skutkovej podstaty trestného činu 3.2 Fakultatívne znaky skutkovej podstaty trestného činu 3.3 Objekt skutkovej podstaty trestného činu 3.3.1 Triedenie objektov 3.3.2 Hmotný predmet útoku 3.4 Objektívna stránka skutkovej podstaty trestného činu 3.5 Subjekt skutkovej podstaty trestného činu 3.5.1 [«2] Pojem	
2. odsek :	spoľahlivosť 100%
[2 »] zákona. „Jednotlivé znaky trestných činov i súhrn týchto znakov sú odrazom skutočnosti, výskytu určitých typických konaní, pred ktorými treba spoločnosť chrániť prostredníctvom sankcií ustanovených na základe Trestného práva. Základom tvorby znakov trestného činu je výskyt určitých konaní, ktoré sú spoločnosťou neakceptovateľné a signalizujú potrebu ich kriminalizácie.“ Úlohou zákonodarcu je vytváranie zákonných znakov tých konaní, pre ktoré je charakteristický ich negatívny spoločenský obsah ako [«2] škodenie	
3. odsek :	spoľahlivosť 96%
[3 »] čin. Osobitá časť Trestného zákona obsahuje nasledovné všeobecné [2 »] atribúty. Sú konkretizované vo forme skutkových podstatí jednotlivých trestných činov. „Súhrn typových znakov trestného činu určitého druhu sa v náuke označuje ako skutková podstata trestného činu. Pojem skutkovej podstaty trestného činu možno definovať ako súhrn typových znakov, ktorými sa od seba odlišujú rôzne druhy trestných činov. Niekedy býva skutková podstata definovaná ako súhrn objektívnych a subjektívnych znakov, ktoré určujú jednotlivé druhy trestných činov a odlišujú ich navzájom. Pod skutkovou podstatou tiež rozumieme súhrn znakov, ktoré charakterizujú určité ľudské správanie ako konkrétny trestný čin. Znaky skutkovej podstaty trestného činu sú uvedené v Trestnom zákone a to ako aj v časti osobitej, tak aj v časti všeobecnej, ktoré tvoria nerozlučnú jednotu“. Skutková podstata trestného činu realizuje ústavnú zásadu, ktorá je, že len zákon [«2,3] stanoví,	
4. odsek :	spoľahlivosť 96%
[2 »] atribúty súčasťou niektorých skutkových podstatí. Vtedy sa z inak nepovinných znakov stávajú znaky povinné, alebo aj obligatórne, teda zákonné. 3.3 Objekt skutkovej podstaty trestného činu Objekt je obligatónym znakom skutkovej podstaty trestného činu. Pod pojmom objekt skutkovej podstaty trestného činu si [«2] môžeme	
5. odsek :	spoľahlivosť 100%
[3 »] čin. „Zmena v kategórii hmotného predmetu útoku v podobe jeho poškodenia sa nazýva účinok.“ Taktiež ako hmotný predmet útoku tak aj účinok je fakultatívnym atribútom objektívnej stránky skutkovej podstaty trestného činu. 3.4 Objektívna stránka skutkovej podstaty trestného činu Objektívna stránka skutkovej podstaty trestného činu je [«3] vonkajšia	
6. odsek :	spoľahlivosť 90%
[4 »] trestného činu tento zákon vyžaduje osobitnú vlastnosť, spôsobilosť alebo postavenie páchatela, páchatelom alebo spolupáchatelom trestného činu môže byť len osoba, ktorá má požadovanú vlastnosť, spôsobilosť alebo postavenie. Účastníkom pri páchaní trestného činu môže byť aj osoba, ktorá nemá požadovanú vlastnosť, spôsobilosť alebo postavenie. Ak zákon ustanovuje, že páchatel musí byť nositeľom osobitnej vlastnosti, spôsobilosti alebo postavenia stačí, ak na túto vlastnosť, spôsobilosť alebo postavenie spĺňa právnická osoba, v ktorej mene páchatel koná.“ 3.6 [«4] Subjektívna	
7. odsek :	spoľahlivosť 84%
[2,4 »] čin je spáchaný z nebanlivosti, ak páchatel a) vedel, že môže spôsobom uvedeným v tomto zákone porušiť alebo ohroziť záujem chránený týmto zákonom, ale bez primeraných alebo ohrozenie nespôsobí, alebo b) nevedel, že svojím konaním môže také porušenie alebo ohrozenie spôsobiť, hoci o tom vzhľadom na okolnosti a na svoje osobné pomery vedieť mal a mohol.“ [«2,4] Zavinenie	



Zarážka zobrazujúca rozsah podobnosti textov.
Čísla v zátvorke zodpovedajú číslu nájdeného dokumentu.



10B09D204743411D86B740B2A82C3691