Nový algoritmus a protokol Theses

Ing. Dita Henek Dlabolová, Ph.D.

2022

MENDELU
Provozně
ekonomická
fakulta

Obsah

1 Oficiální informace Theses k novému algoritmu a protokolu

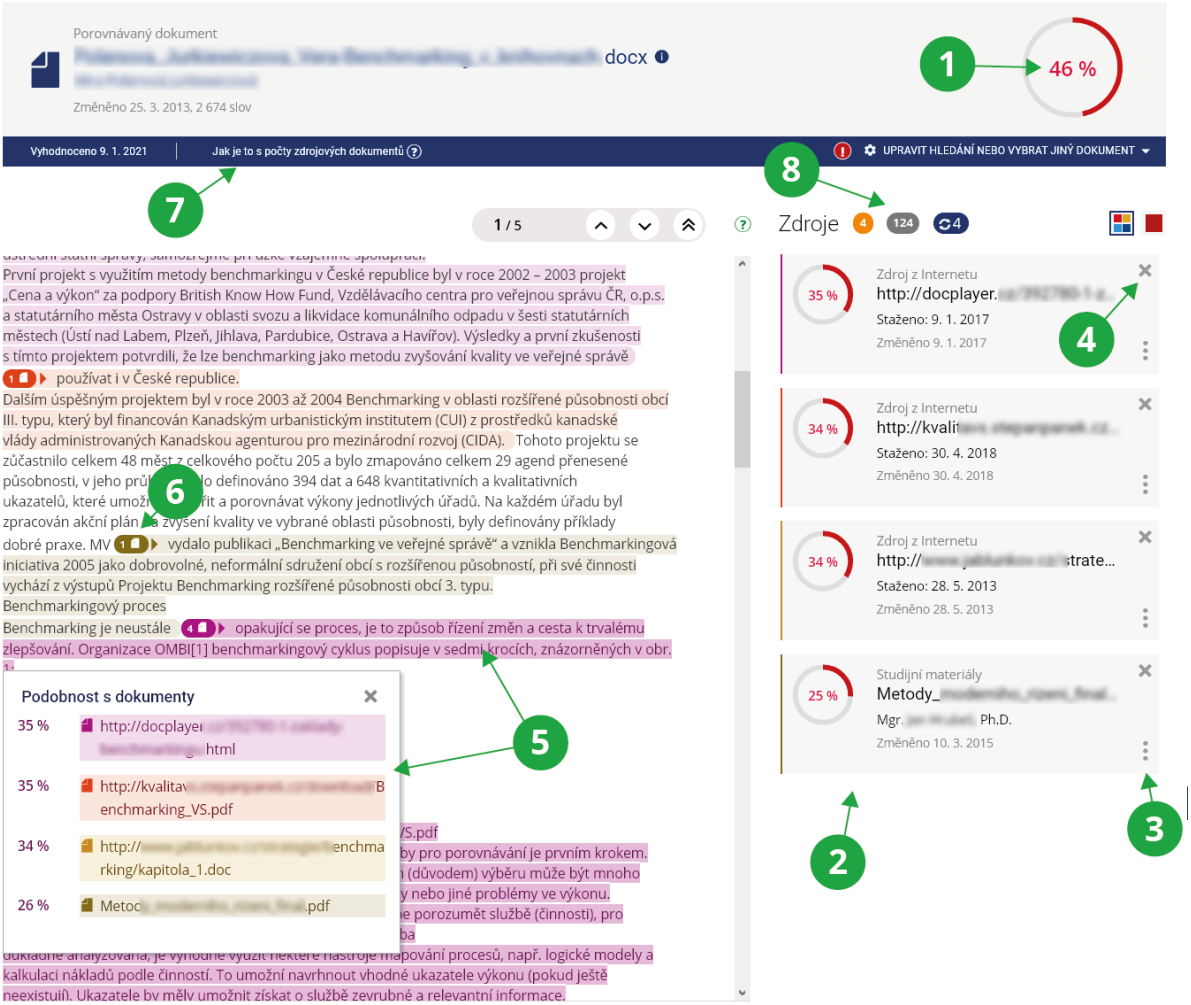
Tým Informačního systému MU, Theses.cz a Odevzdej.cz vyvinul nový algoritmus vyhledávání podobností, který v systémech Theses a Odevzdej.cz. Nový algoritmus vyhledávání podobností lépe odhalí parafrázované (přeformulované) texty, a navíc poskytne i nové funkce, modernější design a přehlednější způsob zobrazení nalezených podobností.

Nápověda k online protokolu je k dispozici zde:
https://theses.cz/napoveda/theses/podobnost#novy_algoritmus

1.1 Informace k novému algoritmu:

- Nový algoritmus se zaměřuje zejména na delší podobné pasáže v textech s tím, že je schopen rozpoznat i poměrně velkou míru přeformulování a parafrázování té stejné informace.
- Naopak nový algoritmus nehledá velmi krátké společné úseky textu o délce jen několika málo slov. Takovéto úseky často bývají různé všeobecně používané definice, zaužívané víceslovné odborné termíny, předepsaná prohlášení v závěrečných pracích a podobné druhy textu, jejichž výskyt ve více dokumentech ještě neznamená plagiátorství.
- Další vlastností nového algoritmu je, že při zobrazování výsledků „přeskakuje“ ty nalezené zdroje, které nepředstavují vůči již zobrazeným žádnou novou přidanou hodnotu a v seznamu nalezených zdrojových dokumentů nefigurují, protože by se už duplikovaly. Tyto "přeskočené" zdroje si lze však zobrazit na kliknutí.
- Podobnosti s jednotlivými zdrojovými dokumenty nyní naleznete vyznačené různými barvami, proti dosavadnímu zvýraznění pouze červeným písmem. Ikonka na začátku barevně zvýrazněné pasáže textu Vám ukazuje, kolik zdrojových dokumentů je s tímto textem podobných, a po jejím rozkliknutí můžete vidět, které zdroje to konkrétně jsou. Kliknutím na vybraný zdroj se zvýrazní podobnosti právě s tímto zdrojovým dokumentem.
- Procento nalezených podobností neurčuje, zda je práce plagiátem, či nikoliv, pouze upozorňuje na určité podezření. Každou práci je nutné posoudit individuálně, zkontrolovat práci nebo správnost citací vždy člověkem, a to odborným pracovníkem v oboru. Podstatné může být zejména to, co konkrétně sděluje nalezená podobná pasáž textu: jestli jde o korektně citovaný převzatý text v přehledové části textu, anebo naopak jestli se jedná o stěžejní část práce, kterou autor prohlašuje za svůj hlavní přínos.

1.2 Jak rozumět online interaktivnímu protokolu na serveru Theses.cz



1. Procento celkové podobnosti s dokumenty v databázi a zdroji z internetu.
2. Seznam zdrojových dokumentů, se kterými je dokument podobný. U každého je zobrazeno procento podobnosti.
3. U každého zdrojového dokumentu je menu pod ikonou tří teček, kde lze o dokumentu zjistit více informací.
4. Za pomoci křížku lze odstranit z výpočtu zdrojový dokument, který není pro porovnání podobností relevantní (například z něj měl student čerpat a má jej řádně citovaný).
5. Po kliknutí na vybranou zvýrazněnou podobnou pasáž se zobrazí dokumenty, se kterými je text podobný.
6. Číslo v oválu označuje počet dokumentů, se kterými je následující pasáž podobná.
7. Po kliknutí se přehledně zobrazí počty zobrazených, přeskočených i vyřazených dokumentů včetně vysvětlení.
8. Ovály s čísly udržují přehled o zdrojových dokumentech. Přeskočené dokumenty lze zobrazit, vyřazené dokumenty obnovit.

2 Vyhodnocování protokolu

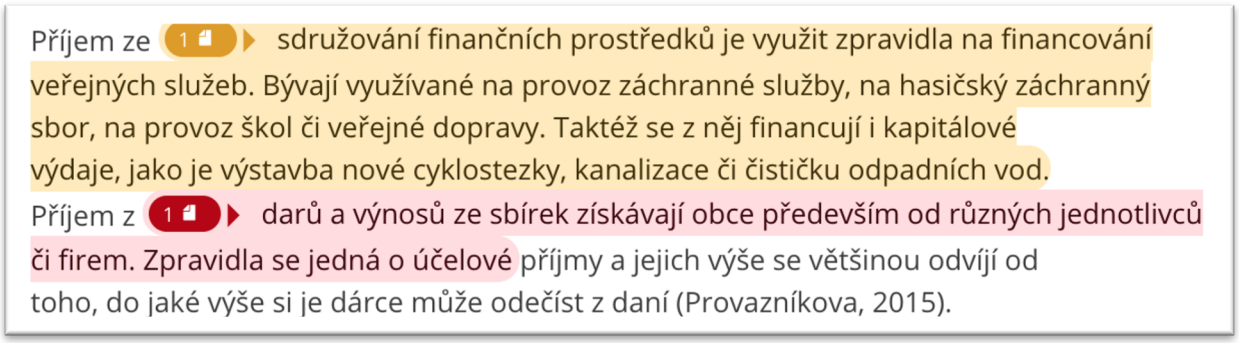
Cílem snažení je především odhalit závažné plagiátorství, tedy úmyslné plagiátorství velkého rozsahu, dle směrnice rektorky 2/2021: *„Za plagiátorství velkého rozsahu je považováno takové plagiátorství, kde převzatá část díla tvoří podstatnou část práce, nebo délka souvislé převzaté části přesahuje jednu normalizovanou stranu“.*

Celkové procento shody nijak nevypovídá o tom, zda je práce plagiát. Pro rozhodnutí je potřeba zkontrolovat vyznačené podobné pasáže. Vyznačené podobnosti představují strojově vyhodnocené parafráze, je potřeba individuálně posoudit povahu těchto pasáží a zda jsou u nich řádně uvedené zdroje.

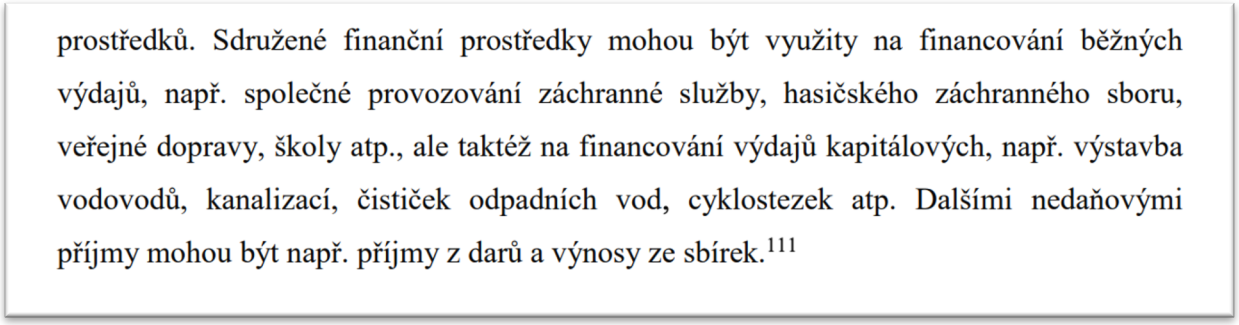
Ideální postup by měl být takový, že vedoucí práce projde jednotlivé vyznačené podobnosti, srovná je s nalezenými dokumenty, a následně rozhodne, zda se jedná o plagiáty, nebo že je vše v pořádku.

2.1 Příklad nalezené a vyhodnocené podobnosti

Podobnost nalezená v závěrečné práci – zkoumáme nyní žlutě vyznačenou podobnost:


Příjem ze 1 sdružování finančních prostředků je využit zpravidla na financování veřejných služeb. Bývají využívané na provoz záchranné služby, na hasičský záchranný sbor, na provoz škol či veřejné dopravy. Taktéž se z něj financují i kapitálové výdaje, jako je výstavba nové cyklostezky, kanalizace či čističku odpadních vod. Příjem z 1 darů a výnosů ze sbírek získávají obce především od různých jednotlivců či firem. Zpravidla se jedná o účelové příjmy a jejich výše se většinou odvíjí od toho, do jaké výše si je dárce může odečíst z daní (Provazníkova, 2015).

Podobná pasáž v nalezeném dokumentu (závěrečná práce z jiné univerzity) a příslušná poznámka pod čarou:


prostředků. Sdružené finanční prostředky mohou být využity na financování běžných výdajů, např. společné provozování záchranné služby, hasičského záchranného sboru, veřejné dopravy, školy atp., ale taktéž na financování výdajů kapitálových, např. výstavba vodovodů, kanalizací, čističek odpadních vod, cyklostezek atp. Dalšími nedaňovými příjmy mohou být např. příjmy z darů a výnosy ze sbírek.[111]

[111] Provazníková, R.: *Financování měst, obcí a regionů, teorie a praxe,* s. 92

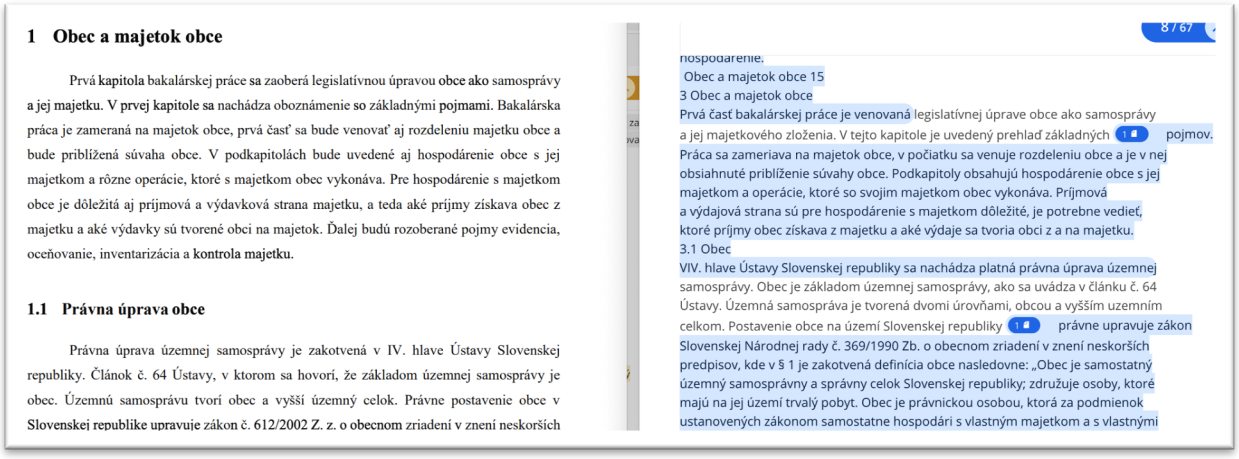
Stejná pasáž v knize, na kterou se odkazují obě závěrečné práce – Provazníková, 2015: *Financování měst, obcí a regionů*, s. 92 (kniha není dostupná online a neměla jsem přístup k celému textu, proto je níže uvedená pasáž neúplná):

> Zejména u malých obcí se ze sdružených prostředků financují veřejní služby, na jejich oddělené financování by obce neměly dostatek finančních prostředků. (…) Sdružené financování vede k využívání úspor z rozsahu. (…) Mohou být použity na financování jak běžných (např. společné provozování záchranné služby, hasičského záchranného sboru, veřejné dopravy, školy), tak kapitálových výdajů (např. výstavba cyklostezek, vodovodů, kanalizace, (…).

Obě práce tedy čerpaly ze stejného zdroje, který řádně odkázaly, tato podobnost je tedy v pořádku.

2.2 Příklad pravděpodobného plagiátu

Theses našel podobnost 21 % odevzdané práce s jiným dokumentem. Při zobrazení podobností pouze s tímto dokumentem, byly podobnosti napříč celou prací. Podobný dokument byl závěrečná práce z jiné školy, která je veřejně dostupná, nebyl tedy problém si ji stáhnout a oba dokumenty porovnat vedle sebe.



1 Obec a majetok obce

Prvá kapitola bakalárskej práce sa zaoberá legislatívnou úpravou obce ako samosprávy a jej majetku. V prvej kapitole sa nachádza oboznámenie so základnými pojmami. Bakalárska práca je zameraná na majetok obce, prvá časť sa bude venovať aj rozdeleniu majetku obce a bude priblížená súvaha obce. V podkapitolách bude uvedené aj hospodárenie obce s jej majetkom a rôzne operácie, ktoré s majetkom obec vykonáva. Pre hospodárenie s majetkom obce je dôležitá aj príjmová a výdavková strana majetku, a teda aké príjmy získava obec z majetku a aké výdavky sú tvorené obci na majetok. Ďalej budú rozoberané pojmy evidencia, oceňovanie, inventarizácia a kontrola majetku.

1.1 Právna úprava obce

Právna úprava územnej samosprávy je zakotvená v IV. hlave Ústavy Slovenskej republiky. Článok č. 64 Ústavy, v ktorom sa hovorí, že základom územnej samosprávy je obec. Územnú samosprávu tvorí obec a vyšší územný celok. Právne postavenie obce v Slovenskej republike upravuje zákon č. 612/2002 Z. z. o obecnom zriadení v znení neskorších

8767

hospodárenie.
Obec a majetok obce 15
3 Obec a majetok obce
Prvá časť bakalárskej práce je venovaná legislatívnej úprave obce ako samosprávy a jej majetkového zloženia. V tejto kapitole je uvedený prehlaď základných pojmov. Práca sa zameriava na majetok obce, v počiatku sa venuje rozdeleniu obce a je v nej obsiahnuté priblíženie súvahy obce. Podkapitoly obsahujú hospodárenie obce s jej majetkom a operácie, ktoré so svojim majetkom obec vykonáva. Príjmová a výdajová strana sú pre hospodárenie s majetkom dôležité, je potrebne vedieť, ktoré príjmy obec získava z majetku a aké výdaje sa tvoria obci z a na majetku.
3.1 Obec
VIV. hlave Ústavy Slovenskej republiky sa nachádza platná právna úprava územnej samosprávy. Obec je základom územnej samosprávy, ako sa uvádza v článku č. 64 Ústavy. Územná samospráva je tvorená dvomi úrovňami, obcou a vyšším územním celkom. Postavenie obce na území Slovenskej republiky právne upravuje zákon Slovenskej Národnej rady č. 369/1990 Zb. o obecnom zriadení v znení neskorších predpisov, kde v § 1 je zakotvená definícia obce nasledovne: „Obec je samostatný územný samosprávny a správny celok Slovenskej republiky; združuje osoby, ktoré majú na jej území trvalý pobyt. Obec je právnickou osobou, ktorá za podmienok ustanovených zákonom samostatne hospodári s vlastným majetkom a s vlastnými

Obrázek 1: Porovnání odevzdané závěrečné práce s vyznačenými shodami (vlevo) a podobného dokumentu (vpravo)

Z textu obou prací je vidět, že podobnosti jsou nad rámec toho, že se jedná o práce na podobné téma, nebo toho, že čerpají ze stejných zdrojů. Shody lze vidět i v nevyznačených částech.

Nejedená se jen o tuto pasáž, obdobnou „inspiraci" starší závěrečnou prací lze najít napříč celou prací. S největší pravděpodobností se jedná o plagiát, navíc velkého rozsahu

2.3 Tipy a triky pro vyhodnocení protokolu

Výše uvedený ideální postup je v praxi pro všechny nalezené podobnosti nerealistický. Níže uvádím několik tipů, jak procházení protokolu urychlit a hlavně, jak odlišit „neškodné" podobnosti od potenciálních plagiátů.

Následující text není oficiálním návodem na vyhodnocení protokolu, jedná se pouze o shrnutí osobních zkušeností a postřehů.

1. Vyhodnocujte protokol v online prostředí Theses

PDF dostupné v UIS je velmi omezené a je prakticky nemožné s ním efektivně pracovat. V PDF pouze klikněte na odkaz „Informace o souboru" na první straně. Odkaz vede do systému Theses, přihlaste se přes EduID a Shibboleth, dostanete se k interaktivní verzi protokolu, která je mnohem přehlednější než PDF.

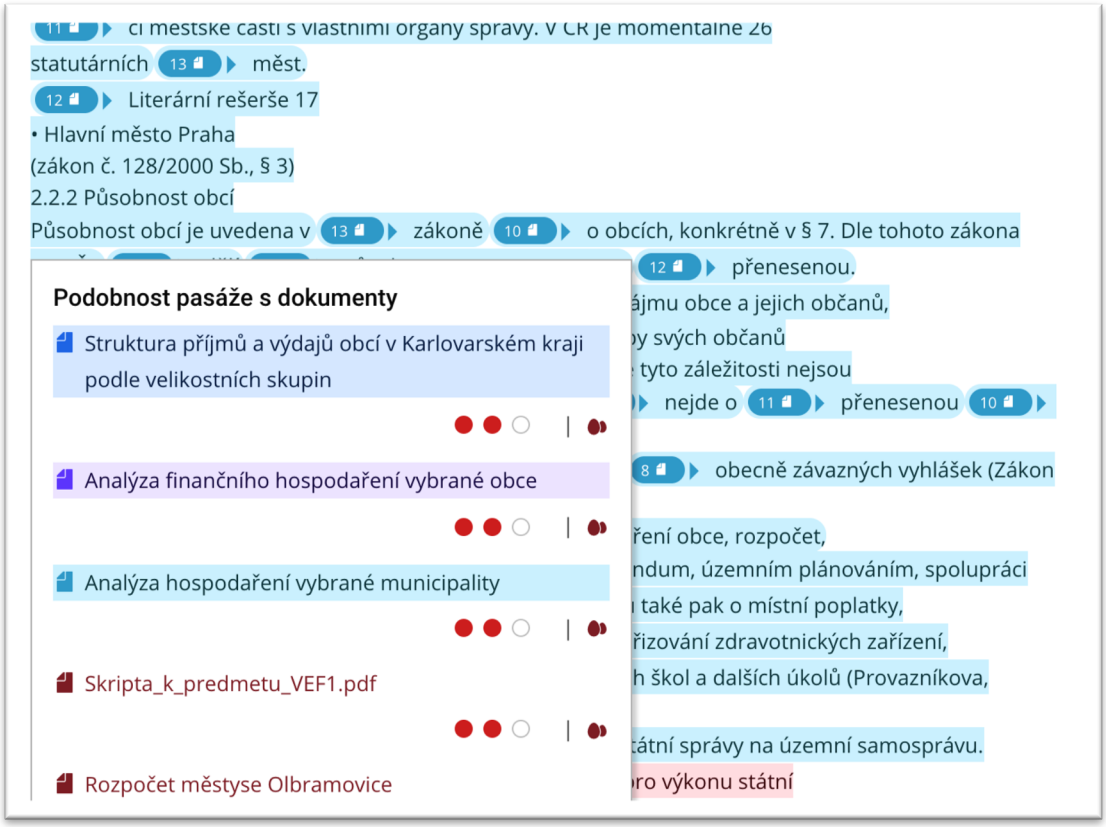
S orientací v online protokolu Vám může pomoci návod Jak rozumět online interaktivnímu protokolu na serveru Theses.cz (na začátku tohoto dokumentu).

2. Přeskočte shody v prohlášení, poděkování, atd.

Přestože Theses tvrdí, že s novým algoritmem už nebudeme upozorňováni stále dokola na shody v povinném prohlášení, není tomu tak. Nalezené podobnosti v „nevýznamových" částech práce tedy můžete přeskočit.

3. Vysoký počet podobných dokumentů indikuje obecné informace

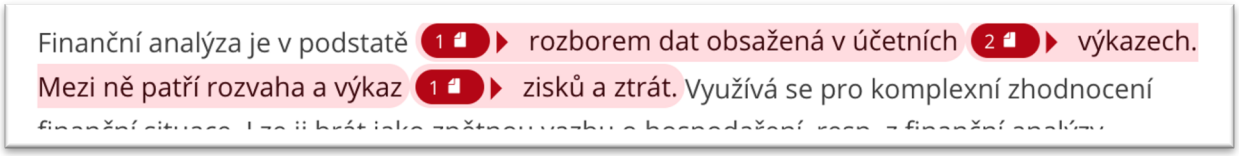
Na začátku každé podobné pasáže je uvedený počet dokumentů, které jsou podobné. Pokud je číslo vysoké (okolo deseti), jedná se pravděpodobně o pasáž, která se vyskytuje v mnoha dokumentech. Z toho lze usuzovat, že se jedná o často opakované nebo obecné informace. Zkontrolujte, že je u vyznačeného textu uvedený zdroj, a dále této části není nutné věnovat pozornost. S největší pravděpodobností se o plagiát nejedná.



Obrázek 2: Příklad podobnosti s mnoha dokumenty. Po přečtení textu lze odhadnout, že se jedná o informace, které se opakují poměrně často. V textu jsou uvedené odkazy na zdroje, nejedná se tedy o plagiát.

4. Velmi krátké vyznačené pasáže jsou pravděpodobně v pořádku

Přestože by nový algoritmus neměl vyznačovat velmi krátké pasáže, v protokolech na ně možná narazíte. S největší pravděpodobností nepředstavují žádný problém.

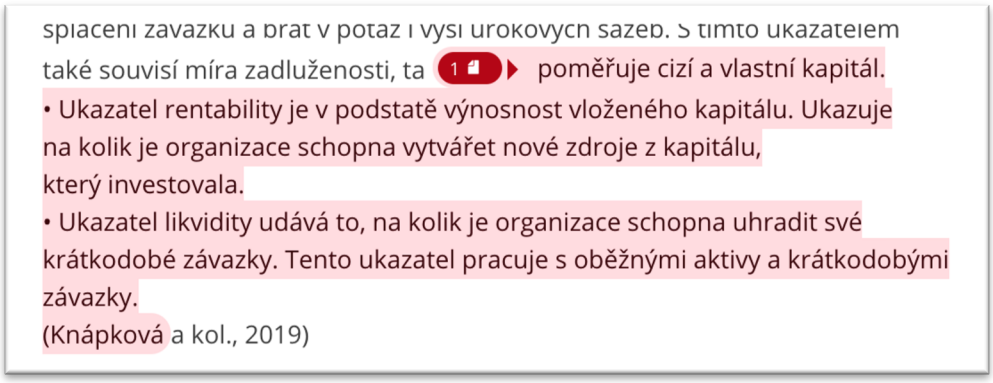


Obrázek 3: Příklad velmi krátkých podobností, navíc s poměrně obecným obsahem. Pokud bychom zkoumali nalezené zdroje, zjistíme, že první část je podobná se zdrojem A, druhá se zdroji A a B, a třetí se zdrojem B. Tuto část tedy můžeme považovat za náhodnou shodu, nejedná se o plagiát.

5. Vyznačené pasáže s uvedeným odkazem na zdroj, pravděpodobně nejsou plagiát

To, že je uvedený odkaz na zdroj, sice nemusí znamenat, že text není plagiát – student totiž mohl text převzít včetně původního odkazu na zdroj. S novým algoritmem je ale bohužel prakticky nemožné dohledat skutečně původní zdroj, protože nevíme, jak moc byl text

parafrázován. Pokud je tedy vyznačený úsek krátký a k informaci přísluší odkaz na zdroj, pravděpodobně se o plagiát nejedná.


splacení závazku a brát v potaz i výši úrokových sazeb. S tímto ukazatelem také souvisí míra zadluženosti, ta 1 poměřuje cizí a vlastní kapitál.
• Ukazatel rentability je v podstatě výnosnost vloženého kapitálu. Ukazuje na kolik je organizace schopna vytvářet nové zdroje z kapitálu, který investovala.
• Ukazatel likvidity udává to, na kolik je organizace schopna uhradit své krátkodobé závazky. Tento ukazatel pracuje s oběžnými aktivy a krátkodobými závazky.
(Knápková a kol., 2019)

Obrázek 4: Příklad krátké podobné pasáže s uvedeným zdrojem – pravděpodobně se nejedná o plagiát.

6. Červeně vyznačené shody jsou méně „nebezpečné“

Prvních 10 dokumentů podobných práci je v textu vyznačených různými barvami. Od jedenáctého dokumentu dál jsou shody vyznačené červeně. S červeným dokumentem je tedy podobnost jen velmi malá a pravděpodobně jsou tyto shody v pořádku.

7. Věnujte pozornost delším podobnostem s jedním zdrojem

Potenciální plagiát může indikovat delší pasáž, která se shoduje s jedním (možná i se dvěma nebo třemi zdroji). Za „delší pasáž“ lze považovat úsek o rozsahu půl strany a více. Takové pasáži věnujte pozornost.

V případě, že se nejedná o řádně ocitovanou parafrázi z literatury, zákona nebo vyhlášky v rešeršní části (což by pravděpodobně bylo v pořádku), prozkoumejte i nalezený dokument (viz níže).

8. U nejvíce shodných dokumentů si zobrazte shody jen s tímto dokumentem

Podobné dokumenty v seznamu napravo jsou seřazené podle míry podobnosti. Podle vyznačené míry shody si u několika prvních zobrazte pouze podobnost mezi prací a tímto dokumentem (u dokumentu vpravo klikněte na tři tečky, a pak „Zobrazit podobnosti pouze s tímto dokumentem“). Pozornost si určitě zaslouží dokumenty s podobností vyšší než 5 %.

Vyznačené podobnosti „prolistujte“ a posuďte podle povahy vyznačených podobností. Pokud se jedná o řádně odkázané parafráze nebo obecné formulace, pravděpodobně je vše v pořádku.

Pokud podobnosti vypadají podezřele, prozkoumejte i nalezený dokument (viz níže).

9. Prozkoumání dokumentu při podezřelé shodě

Tím, že Theses vyznačuje všechny podobnosti, o plagiátu lze rozhodnout, až když se podíváme i do nalezeného podobného dokumentu.

Dostupný dokument

K části nalezených dokumentů se lze dostat jedním z následujících způsobů:

- Tři tečky v seznamu dokumentů otevřou detaily podobného dokumentu, pak klikněte na název dokumentu. Toto Vás pravděpodobně přivede na portál závěrečných prací dané instituce, kde bude možné stáhnout si PDF s daným dokumentem.
- Ikonka zeměkoule je přímý odkaz na dokument.

V dokumentu je pak potřeba najít onu podobnou pasáž, což může být skutečně detektivní práce. Doporučuji vyhledat v podobném dokumentu unikátní slova nebo fráze, která by se těžko parafrázovala (v příkladu v části 2.1 o financování veřejných služeb pomohlo slovo „cyklostezka“).

Když se Vám podobnou sekci podaří najít, je na Vašem posouzení, zda se jedná o podobnost, která je v pořádku, nebo zda student opisoval.

Při posuzování je dobré zobrazit si dokumenty vedle sebe a postupně je oba projít. Pozornost si zaslouží i text okolo vyznačených shod. Indikátorem závažného plagiátorství je obdobná struktura obou dokumentů, použití stejných zdrojů, podezření pomůže potvrdit i shodný text, je jen „omáčkou“ (např. „v první části je popsáno téma X, následuje část věnovaná tématu Y“). Toto je vidět v příkladu v části 2.2.

Nedostupný dokument

K některým nalezeným dokumentům se bohužel výše uvedenými způsoby nedostaneme. U takových podobností si můžete pomoci Googlem – je možné zadat do vyhledávače několik frází, které by mohly být unikátní pro daný text. Pokud ani Google nepomůže, spolehněte se prosím na svůj cit a pedagogickou zkušenost. Posuzujte podle rozsahu a významu pro práci.

3 Závěrem

I s novým protokolem a algoritmem naslouchejte především svému pedagogickému citu a zkušenosti. Řiďte se i vztahem s konkrétním studentem – pokud pravidelně konzultoval a upravoval text podle Vašich připomínek, je větší pravděpodobnost, že nalezené podobnosti jsou v pořádku.

Cílem systémů pro podporu odhalování plagiátorství není „hon na čarodějnice“. Nechceme perzekuovat studenty kvůli jedné větě, kterou zapomněli ocitovat. Systémy nám pomáhají odhalit studenty, kteří neodvedli řádně svou práci a neprokázali tak svou závěrečnou prací, že mají znalosti a schopnosti, které by měli mít.

Pokud máte vážné podezření na plagiát, postupujte dle platné směrnice rektorky (2/2021 – Prevence a odhalování plagiátů). V případě závažného plagiátorství za závěrečnou práci neudělíte zápočet a podáte podnět k zahájení disciplinárního řízení. V případě pochybností se na mě neváhejte obrátit, v kontaktním centru je také možnost požádat o kontrolu systémem PlagScan.