Vysoká škola ekonomická v Praze

Fakulta managementu

BAKALÁRSKA PRÁCA

2014 Tomáš Misák

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

FAKULTA MANAGEMENTU V JINDŘICHOVĚ HRADCI

KATEDRA EXAKTNÍCH METOD



AKTUÁLNE MOŽNOSTI ODHAĽOVANIA PLAGIÁTORSTVA V TEXTOCH

Autor bakalárskej práce: Tomáš Misák

Vedúci bakalárskej práce: Ing. Jiří Přibil, Ph.D.

Rok obhajoby: 2014

Čestné prehlásenie

Prehlasujem, že som bakalársku prácu na tému „Aktuálne možnosti odhaľovania plagiátorstva v textoch“ vypracoval samostatne a všetku použitú literatúru a ďalšie pramene som riadne označil a uviedol v priloženom zozname.

V Jindřichovom Hradci dňa 20. 5. 2014

.....
podpis autora

Zadanie práce

(papier obdržíte v škole)

Anotácia

Bakalárska práca na tému „Aktuálne možnosti odhaľovania plagiátorstva v textoch“ je zameraná na plagiátorstvo v záverečných prácach študentov. Objasňuje pojem plagiátorstvo, čo to vlastne je a vysvetľuje ďalšie pojmy ktoré sú s ním spojené. Popisuje rôzne typy a dôvody prečo vôbec študenti plagiujú. Zaoberá sa aj citáciami a v menšej časti sa venuje aj autorskému zákonu ktorý je s týmto spojený.

Teoretická časť práce je zameraná na metódy odhaľovania plagiátorstva kde je určitý text poskytnutý viacerým študentom za účelom aby vytvorili plagiát. Ten je potom otestovaný v programe Theses ktorý je navrhnutý za účelom odhaľovania plagiátov. Výsledkom tohto pokusu je otestovanie funkčnosti programu a jeho celkové správanie počas detekcie plagiátu.

Annotation

Bachelor thesis on theme „Current possibilities of detecting plagiarism in texts“ is focused on plagiarism in the theses of students. Explains the concept of plagiarism, what it actually is and explains other concepts that are associated with it. Describes different types and reasons why exactly students are plagiaring. It also deals with citations and a smaller part is devoted to Copyright Act that is associated with it.

The theoretical part focuses on methods for the detection of plagiarism where some texts are provided to several students in order to create plagiarism. After that it is tested through program named Theses which is designed for detection of plagiarism. The result of this is testing of functionality and its overall behavior in detection process.

Poďakovanie

Na tomto mieste by som rád poďakoval Ing. Jiřímu Přibilovi, Ph.D. za pomoc a užitočné rady pri písaní bakalárskej práce.

Obsah

I. Úvod ... 1

II. Plagiátorstvo ... 2

II. 1. Definícia ... 2

II. 2. Úmyselný plagiát ... 2

II. 3. Kompilácia ... 3

II. 4. Parafráza ... 3

II. 5. Všeobecne známy fakt ... 3

II. 6. Využívanie vlastných diel ... 4

II. 7. Kryptomnézia (skrytá pamäť) ... 4

III. Porovnanie právnej úpravy v ČR a SR ... 4

III. 1. Autorský zákon ... 4

III. 2. Zákon o vysokých školách ... 5

III. 3. Zhrnutie zákonov ... 5

IV. Druhy plagiátorstva ... 6

IV. 1. Druhy plagiátorstva podľa Dr. C. Barnbauma ... 6

IV. 1.1. Copy & paste ... 6

IV. 1.2. Zámena poradia slov ... 7

IV. 1.3. Zmena štýlu v texte ... 7

IV. 1.4. Plagiátorstvo metafor ... 7

IV. 1.5. Ideové plagiátorstvo ... 8

IV. 2. Ďalšie príklady plagiátorstva ... 8

V. Príčiny výskytu plagiátorstva ... 9

VI. Citácie ... 10

VI. 1. Citácie podľa normy ISO 690 ... 11

VI. 1.1. ISO 690 Dokumentácia - Bibliografické odkazy ... 11

VI. 1.2. ISO 690-2 Informácie a dokumentácia ... 12

VI. 2. Generátor citácií ... 12

VII. Nástroje na odhaľovanie plagiátorstva ... 16

VII. 1. Turnitin ... 16

VII. 1.1. iThenticate ... 17

VII. 2. Copy catch ... 17

VII. 3. Essay verification engine ... 17

VII. 4. WCopyfind 18
VII. 5. Scan my Essay 18
VII. 6. Theses 19
VII. 6.1. Praktické použitie Thesesu 21
VIII. Záver 32

I. Úvod

S plagiátorstvom ako takým sa stretávame nie len v záverečných prácach či dokumentoch, ale aj v rôznych iných smeroch ako napríklad: filmoch, hudbe, móde, architektúre a v ďalších iných. V podstate by sa dalo povedať, že všade možne. Táto práca sa však zameriava len na plagiátorstvo v textových dokumentoch a to hlavne kvôli tomu, že aj ja som študentom a tým pádom sa ma táto téma dosť výrazne týka.

V dnešnej dobe, dalo by sa povedať pokročilej informačnej dobe, je najmä internet jedným z médií, ktoré sa rozvíja veľmi vysokou rýchlosťou a zasahuje do života obrovského počtu ľudí. Kombinácia rýchleho šírenia dostupnosti internetu a pokrok technológií, dáva tomu, že ľuďom sa naskytuje čím ďalej, tým viacej, či už voľne alebo platene dostupných informácií. A tak isto ako sa autori nejakého maliarskeho diela môžu nechať inšpirovať autorom iným, tak isto sa postupne čoraz ľahšie a ľahšie môžu nechať inšpirovať napríklad študenti prácou niekoho iného. V horšom prípade sa nenechávajú len inšpirovať ale doslova kopírujú z cudzích prác do vlastných, bez riadneho citovania alebo aspoň minimálneho spomenutia pôvodného autora. Kvôli tomuto všetkému školy prišli k tomu, že budú záverečné práce kontrolovať. Slúžia im k tomu detekčné nástroje, ktorých existuje viacero, niektoré sú však platené a iné sú na druhú stranu zdarma. Taktiež sa líšia či už v kvalite, alebo napríklad vo formátoch, ktoré dokážu kontrolovať. Táto bakalárska práca ďalej nerieši len základné pojmy, či zákon ale aj funkčnosť a efektivitu konkrétneho systému, ktorý slúži na detekovanie plagiátorstva a to systému Theses.

V teoretickej časti práce najprv vysvetľujem pojem plagiátorstvo ako taký. Ďalej je vysvetlených pár pojmov, ktoré s danou tematikou úzko súvisia a hneď potom porovnávam právnu úpravu v Slovenskej a Českej republike. Konkrétne sa jedná o autorský zákon a zákon o vysokých školách. Následne popisujem rôzne druhy plagiátorstva a samozrejme a dôvody prečo k nemu dochádza. Po dôvodoch plagiátorstva nasledujú citácie, správne citovanie podľa normy ISO 690 a príklady generátorov citácií, ktoré sú dostupné. Na záver teoretickej časti popisujem pár vybraných nástrojov, ktoré slúžia k odhaľovaniu plagiátov.

V praktickej časti budem v podstate skúšať systém Theses. Spôsobom takým, že vytvorím dokument, ktorý poskytnem pár ľuďom aby ho prepísali podľa seba a následne ho nechám systémom overiť, aby som zistil ako vôbec funguje, poprípade aby som zistil jeho nedostatky. Keďže som študent zo Slovenska, tak dokument nechám preložiť aj do češtiny, aby som zistil ako systém porovnáva slovenčinu a češtinu.

II. Plagiátorstvo

II. 1. Definícia

Dobrá otázka na začiatok je, čo je to vlastne plagiátorstvo. Norma ČSN ISO 5127 definuje plagiát ako vydávanie duševného diela iného autora, ktoré sme si zapožičali alebo napodobnili buď celé alebo len z časti, za svoje vlastné. (1)

Nemusí to byť len úmyselné skopírovanie, či ukradnutie textu niekoho iného a jeho následné prezentovanie za svoje vlastné dielo. Ale, môže nastať napríklad aj pri nesprávnom použití citácie alebo nechcené vynechanie citácie niektorého zo zdrojov, z ktorých sme do našej práce čerpali. Alebo sa ho dá dopustiť aj nedostatočnou prácou s originálnym textom, napríklad nedostatočnou parafrázou alebo kompiláciou textu. Plagiátormi nie sú teda len tí ľudia, ktorí takéto texty vytvárajú, či vedome alebo nevedome, ale aj tí, ktorí k plagiátorstvu navádzajú. Čiže napríklad niekto taký, kto ponúka takéto texty či už zadarmo alebo za pár drobných, na tom nezáleží. (2) V tomto objasňovaní je nutné vysvetliť aj pojmy ako úmyselný plagiát, kompilát, parafráza, všeobecne známy fakt, využívanie vlastných diel a „skrytá pamäť“ ktoré sú taktiež spojené s plagiátorstvom.

II. 2. Úmyselný plagiát

Znamená to nasledovné:

- Vyslovene opísať alebo skopírovať výtvor niekoho iného a následne ho prezentovať za svoj vlastný, bez toho aby sme použili patričné citačné prostriedky.
- Prevziať a publikovať cudziu prácu (či už seminárnu alebo inú), aj vrátane tých, ktoré ešte neboli dokončené a riadne odovzdané.
- Vydávanie kompilácie alebo jej časti za vlastný text.
- Skopírovanie grafiky (napríklad obrázok, graf) bez toho, aby sme uviedli odkaz na pôvodný zdroj a bez citácie.
- Skopírovanie názvu, štruktúry (napríklad obsahu, osnovy a iné) cudzieho textu, v extrémoch až do takej miery, že je možná zámena oboch diel.
- Úmyselné neuvádzanie niektorých využitých zdrojov.

- Kúpenie či stiahnutie voľne dostupnej práce a jej následné vydávanie za vlastnú. (2)

II. 3. Kompilácia

Ako už bolo spomenuté čo sa plagiátov týka, je potrebné objasniť aj ďalšie pojmy, ktoré sú s nimi spojené. Ďalším v poradí je teda kompilácia, ktorá je vysvetlená ako akt, či proces zoskupovania vecí dokopy. (3) Čiže v oblasti textových plagiátov to môžeme chápať ako zber jednotlivých častí textu z rôznych zdrojov a ich následné „zlepenie“ dokopy, čím nám vznikne jednotný text. Samozrejme ako v každom inom prípade, je aj tu nutné správne použiť citácie a uviesť zdroje, z ktorých sa čerpalo. Ináč kompilácia nebude nič iné ako ďalší plagiát.

II. 4. Parafráza

Parafráza je preformulovanie celého textu alebo jeho časti vlastnými slovami. Parafráza môže veľmi ľahko vzniknúť napríklad prekladaním z cudzieho jazyka, alebo prevzatím myšlienky iného autora vyloženej vlastnými slovami. Avšak treba brať na vedomie aj to, že parafráza neznamená zámenu pár slov v origináli, napríklad synonymami. Ale musí prísť aj k zmene štýlu, pričom sa zachováva pôvodná myšlienka textu. Taktiež tu je aj jeden extrém, kedy sa nezachová ani len pôvodná myšlienka a v tom prípade sa o parafrázu už nejedná. (4)

II. 5. Všeobecne známy fakt

Jedná sa o informácie ktoré sú verejnosti veľmi dobré známe. V tomto prípade nie je potrebné uvádzať zdroj, pretože informácia je každému veľmi dobre známa. Necitovanie či neuvádzanie zdrojov pri všeobecne známych faktoch sa nedá považovať za prejav plagiátorstva (5). Napríklad ako to, že pes má štyri nohy. Ak by sme ale chceli pokračovať v tom že akej dĺžky majú nohy jednotlivé plemená, tak by sme už zdroj uviesť museli. To už všeobecne známe nie je a zisťovali by sme to určite z niekoho konkrétnej práce či výskumu.

II. 6. Využívanie vlastných diel

Ako už z názvu vyplýva, tak sa jedná o využívanie informácií z vlastných diel. Avšak uvedeniu zdroju je taktiež potrebné aj v tomto prípade. Podaktorí môžu zabudnúť na uvedenie zdrojov práve preto, že čerpajú z vlastného výtvoru. (2)

II. 7. Kryptomnézia (skrytá pamäť)

Kryptomnézia môže veľmi jednoducho predchádzať vzniku plagiátu. Nastáva vtedy, keď vytvoríme niečo, čo síce máme v našich spomienkach v podvedomí, ale nepamätáme si to. Následne vzniká „nová myšlienka“, ktorá je prevzatá od iného autora. Aj keď nevedome, lebo si to nepamätáme, alebo si iba pamätáme že niekde možno niečo podobné už bolo. (6)

III. Porovnanie právnej úpravy v ČR a SR

III. 1. Autorský zákon

Autorský zákon upravuje autorské práva a tie keď sú porušované, tak sa dopúšťame plagiátorstva.

V Slovenskej republike upravuje tieto práva zákon č. 618/2003 Z. z., zákon o autorskom práve a právach súvisiacich s autorským právom (autorský zákon).

V Českej republike ich upravuje zákon č. 121/2000 Sb., zákon o právu autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov (autorský zákon).

V obidvoch krajinách sa však oba zákony skoro zhodujú. Oba považujú za predmet autorského práva to isté. A to napríklad dielo literárne, umelecké, vedecké ktoré sú výsledkami tvorivej činnosti autora.

Autor diela je v zákonoch zadefinovaný ako:

- Autor je fyzická osoba, ktorá dielo vytvorila.

- Autorom súborného diela ako celku je fyzická osoba, ktorá vybrala alebo usporiadala jeho obsah; práva autorov diel takto vybraných alebo usporiadaných v súbornom diele nie sú dotknuté. (7) (8)

III. 2. Zákon o vysokých školách

Plagiátorstvo upravuje ďalej v Českej aj Slovenskej republike zákon o vysokých školách.

A to na Slovensku zákon č. 131/2002 Z.z., ktorý hovorí o nasledovnom: ,,Záverečná práca nesmie neoprávnene zasiahnuť do práv alebo právom chránených záujmov tretích osôb, najmä nesmie porušovať práva duševného vlastníctva tretej osoby alebo neoprávnene nakladať s utajovanými skutočnosťami alebo osobnými údajmi, dôvernými informáciami či obchodným tajomstvo tretej osoby.“ (9)

V Čechách ho upravuje zákon č. 111/1998 Sb. a čo sa týka plagiátorstva hovorí len o zverejňovaní prác. ,,Vysoká škola nevýdělečně zveřejňuje disertační, diplomové, bakalářské a rigorózní práce, u kterých proběhla obhajoba, včetně posudků oponentů a záznamu o průběhu a výsledku obhajoby prostřednictvím databáze kvalifikačních prací, kterou spravuje. Způsob zveřejnění stanoví vnitřní předpis vysoké školy.“ (10)

III. 3. Zhrnutie zákonov

V obidvoch krajinách sú tieto zákony veľmi podobné. Rovnako definujú, kto je autorom a pre obe krajiny je predmetom autorského zákona to isté. Rozdiel nie je ani v zákone o vysokých školách, kde oba hovoria o tom, akým spôsobom sú zverejňované práce študentov a aké majú práva pri ich zverejňovaní.

Ďalej čo sa jedná disciplinárnych priestupkov, v zákone o vysokých školách, ktoré sú taktiež spojené s touto tematikou. Bez rozdielu krajiny sa za disciplinárny priestupok ukladajú študentovi nasledujúce opatrenia: pokarhanie, podmienečné vylúčenie alebo vylúčenie zo štúdia. A platí aj to, že priestupok je premlčaný, ak od jeho spáchania ubehne viac než rok, ale nezapočítava sa tam doba, kedy osoba nie je študentom. Alebo, podľa slovenskej verzie sa tam nezapočítava doba, kedy je prerušené štúdium a ani doba, kedy prebieha konanie o danom priestupku. (9) (10)

Podrobnejšie predpisy či povinnosti sú samozrejme upravené v jednotlivých vnútorných poriadkoch konkrétnych škôl.

IV. Druhy plagiátorstva

Čo sa tejto témy týka, je ďalej vhodné si vymenovať, ako dochádza k plagiátorstvu podľa rôznych zdrojov.

IV. 1. Druhy plagiátorstva podľa Dr. C. Barnbauma

Barnbaum vo svojom príspevku poukazuje na to že väčšina študentov pozná iba CTRL+C, CTRL+V ako druh plagiátorstva. Avšak on ich rozlišuje viacej a to nasledujúce: (11)

- Copy & paste
- Zámena poradia slov
- Zmena štýlu v texte
- Plagiátorstvo metafor
- Ideové plagiátorstvo

IV. 1.1. Copy & paste

Jedná sa o najzákladnejšiu a najjednoduchšiu formu plagiátorstva. Je to vlastne skopírovanie celej myšlienky, alebo textu iného autora a následné vloženie do vlastnej práce. A netreba na to spraviť nič zložitejšie, ako stlačiť dokopy štyri klávesy na klávesnici (CTRL+C, CTRL+V). Alebo taktiež štyri aj na myši (doslova pravé, ľavé, pravé, ľavé). Táto jednoduchosť zabezpečuje asi aj popularitu. Prevenciou je už klasická priama citácia pomocou úvodzoviek. Ako príklad by sa dalo uviesť nasledujúce:

Zdrojový text: Pri tomto športovom aute, ktoré symbolizuje náš nekonvenčný prístup, sme sa inšpirovali japonským pojmom Jinba Ittai - „Kôň a jazdec sú jedno“.

Vytvorený plagiát: Pri tomto športovom aute, ktoré symbolizuje náš nekonvenčný prístup, sme sa inšpirovali japonským pojmom Jinba Ittai - „Kôň a jazdec sú jedno“. (12)

IV. 1.2. Zámena poradia slov

Ako už názov hovorí, ide o veľmi jednoduché zamenenie poradia slov. V tomto prípade sa nejedná o priamu citáciu, pretože vety textu nemajú identickú konštrukciu s originálom. Nasleduje jednoduchý príklad takejto zámeny: (11)

Zdrojový text: Jedinečnosť odtlačku prstu si uvedomovali už v starovekom Grécku.

Vytvorený plagiát: V starovekom Grécku si už uvedomovali jedinečnosť odtlačku prstu. (13)

IV. 1.3. Zmena štýlu v texte

Pri tomto druhu nedochádza k presnému opísaniu či skopírovaniu originálneho textu. Nekopírujú sa identicky celé texty, odstavce či vety. Ale zamieňa sa napríklad čas vo vetách (prítomný za minulý), rod (mužský za stredný) alebo sa používajú rôzne synonymá originálnych slov. Tento štýl sa samozrejme používa väčšinou kombinovaním hore uvedených. (11) Príklad:

Zdrojový text: Teraz keď sa za mnou ženú a bez oddychu, počujem ich kroky hlasnejšie. Prenasledujú ma už roky.

Vytvorený plagiát: Momentálne ako ma naháňajú a bez prestávky, čuť stupaje intenzívnejšie. Sú mi v pätách už neskutočne dlho. (14)

IV. 1.4. Plagiátorstvo metafor

Každý autor používa vo svojich výtvoroch vlastné metafory. Sú súčasťou autorovho osobitného štýlu. Takže následné odhalenie nie je až také zložité, pretože aj keď to zväčša môže byť len pár slov, tak aj tak vybočujú zo štýlu autora, ktorý sa dopúšťa plagiátorstva. Pri ich použití sa doporučuje použitie úvodzoviek a následné priradenie metafory k patričnému autorovi. Z dôvodu, aby nebol nikto okrádaný o svoju „originalitu a kreativitu“. Opäť krátky príklad:

Zdrojový text: V infračervenom žiarení môžeme vidieť trblietavú šperkovnicu novozrodených hviezd, ako vykúkajú z oblakov prachu, ktoré ležia medzi nami a stredom našej galaxie.

Vytvorený plagiát: Títo novorodenci sa trblietajú ako šperkovnica a zdá sa, že na nás vykúkajú spoza prachu, ktorý ich zastiera. (11)

IV. 1.5. Ideové plagiátorstvo

Ideové alebo plagiátorstvo myšlienok je vlastne prezentovanie myšlienky iného autora za svoju vlastnú. K tomuto väčšinou dochádza vtedy, keď nevieme rozlíšiť či to už niekto niekedy spomenul, alebo či je to naozaj originálna myšlienka vytvorená našim mozgom. V takomto prípade môže ako prevencia poslúžiť jednoduché spýtanie sa kompetentných v obore. Príklad na záver:

Zdrojový text: Horúce hviezdy pri teplote 30 000 stupňoch vyžarujú viacej modrého svetla ako červeného.

Vytvorený plagiát: Hviezdy ktoré sa považujú za horúce majú 30 000 stupňov. (11)

IV. 2. Ďalšie príklady plagiátorstva

Samozrejme existuje viacej druhov plagiátorstva. A každý zdroj môže uvádzať rozličné, alebo rovnaké len pod iným názvom. Plagiarism.org (15) uvádza na svojich stránkach nasledujúce možnosti, ako sa dá postupovať pri vytváraní plagiátu (nasledujúce názvy budú uvedené v angličtine z dôvodu nevhodnosti prekladu):

- ,,Clone“ - skopírovanie slova po slove ako takých bez akejkoľvek zmeny
- ,,CTRL - C“ - dalo by sa povedať že je to taký slabý odvar prvej možnosti. Nejedná sa totiž o skopírovanie celého textu, ale len jeho častí. Avšak, taktiež bez zmien.
- „Find - Replace“ - hrubá podstata práce je v tomto prípade zachovaná. Zmenené sú len tie najdôležitejšie slová a použité frázy v práci.
- „Remix“ - ako už pomenovanie tohto druhu naznačuje, jedná sa o „zlepenie“ textu z viacerých zdrojov do jedného.
- „Recycle“ - použité sú tu myšlienky z práce ktorú už autor v minulosti vytvoril. Zabúda sa však na citácie. Či už nechceným omylom alebo za účelom.
- „Hybrid“ - je kombináciu citovaných častí v práci, tak ako sa predpokladá a skopírovaných častí. Účelom chce byť zrejme zamaskovanie prevzatých myšlienok.
- „Mashup“ - jedná sa o mix, alebo kombináciu skopírovaných myšlienok z viacerých zdrojov a ich následné zlepenie do jednotného textu.

- „404 Error“ - citácie tu sú síce použité tak ako sa očakáva. Ale, to čo je v nich citované buď vôbec neexistuje, alebo sú to nepresné informácie k danej problematike.

- „Aggregator“ - obsahuje správne citácie, aj existujúce, nuž v práci sa nevyskytuje skoro žiadna autorova originálna myšlienka.

- „Re - tweet“ - citácie sú opäť v úplnom poriadku ale autor sa až príliš drží formulácie originálneho textu alebo štruktúry. Poprípade oboch.

V. Príčiny výskytu plagiátorstva

Prečo sa vôbec študenti dopúšťajú plagiátorstva ? Je to veľmi dobrá otázka, ale odpoveď je na druhú stranu veľmi ťažká. Existuje mnoho dôvodov, prečo tomu tak je a na každého jedinca sa môže vzťahovať iný. Niekto môže byť príliš lenivý na to, aby obetoval svojej práci viacej času, tak volí automaticky tú ľahšiu cestu. Niekto ďalší má možno aj iné povinnosti a nachádza sa v časovej tiesni. Na niekoho možno tlačí škola, ale nájde sa aj viacero ďalších dôvodov prečo sa to deje. Príčiny plagiátorstva môžu byť teda nasledujúce (odlišné zdroje môžu kľudne uvádzať rôzne ďalšie):

- Nedostatok vyhľadávacích schopností

  Mnoho autorov trpí nedostatkom schopnosti vyhľadávať. Či už vyhľadávať internetové zdroje, v mestskej či školskej knižnici, alebo v rôznych databázach, ktoré sú verejnosti k dispozícii. Následne to vedie k tomu, že má autor tendenciu kopírovať z cudzích prác.

- Problém hodnotenia internetových zdrojov

  V dnešných časoch, kedy je internet rozšírený už takpovediac po celom svete, je veľmi ťažké s istotou povedať, čo je a čo nie je spoľahlivý zdroj. Prístup na internet sa dá získať už skoro všade a pomaly každý laik tak má možnosť prezentovať svoje myšlienky, ako tie najspoľahlivejšie informácie na internete. Nutné je teda hľadať také zdroje, ktoré sú overené a ktoré nemôžu všetci upravovať.

- Zmätok v riadnom citovaní zdrojov

  V prvom rade je dôležité zachovať jednotný štýl pri citovaní použitých zdrojov v práci. Nie pri každom zdroji odlišný. Avšak, často sa používajú aj on-line zdroje, ktoré sa citujú ťažšie ako knižné zdroje. A to z toho dôvodu,

že adresa stránky, na ktorú odkazuje zdroj, sa môže hocikedy zmeniť alebo URL môže byť príliš dlhé. Kvôli tomu je asi najbezpečnejšie k zdroju pripísať dátum, kedy sme z neho čerpali.

- Tlak zo strany rodiny, školy či iných

  Členovia rodiny alebo školy môžu vyvíjať tlak na študentov tým, že im bude pridelený napríklad nejaký benefit pod podmienkou dobrých známok. To môže fungovať pre niektorých ako spúšťač k tomu, že začnú plagizovať, aby „nesklamali“.

- Nesprávne hospodárenie s časom

  Neefektívne využitie času vedie k tomu, že študenta alebo autora zaplaví množstvo práce, ktorú má napísať. Až kým sa dostane do štádia, kedy zistí, že vlastné sily mu už nestačia k tomu, aby sa zmestil do časového limitu. A to všetko predchádza tomu, že autor vytvorí len ďalší plagiát. (16)

VI. Citácie

Správne citovanie zdrojov je jedným z hlavných bodov každej práce. Ak autor použije cudzí zdroj, tak ho musí aj riadne odcitovať, so všetkými potrebnými náležitosťami. Pri nedodržaní tohto jednoduchého postupu sa autor dopúšťa plagiátorstva. Či už chceného alebo nechceného.

Citácie ako také v texte nenájdeme, iba odkaz ktorý nás presmeruje na iné miesto v práci. Ich kompletnú formu potom nachádzame buď v poznámke pod čiarou, alebo na konci práce v podobe bibliografických odkazov. (17)

S citáciami sú samozrejme spojené aj citačné normy, ktoré by sa pri ich vytváraní mali riadne dodržiavať. V Českej aj Slovenskej republike sa najčastejšie používa citačná norma ISO 690.

ISO je skratka medzinárodnej organizácie pre normalizáciu (International Organization for Standardization). Je najväčšou vývojárskou spoločnosťou, čo sa medzinárodných noriem týka. Tieto normy udávajú rôzne špecifikácie pre či už výrobky alebo služby. Pomáhajúc priemyslu aby bol účinnejší a efektívnejší. Normy sa rozvíjali prostredníctvom celosvetového konsenzu, aby pomáhali prelomiť bariéry v medzinárodnom obchode. Spoločnosť bola založená v roku 1947 a odvtedy dala

vzniku viacej než 19 500 medzinárodným štandardom vo viacerých oblastiach. Taktiež sa postarala aj o už spomínanú normu ISO 690.

ISO 690 je norma ktorá sa snaží napovedať o tom ako sa majú spracovávať bibliografické odkazy a citácie. Je možné ju použiť pre takéto odkazy a citácie vo všetkých druhoch informačných zdrojov vrátane príspevkov, patentov, kartografických materiálov, fotografií, hudby a ďalších iných. Samozrejme nenariaďuje presný štýl, ako by odkazy či citácie mali vyzerať. (18)

Pre každú krajinu má ISO norma špecifické označenie. Napríklad v Českej republike sa označuje ako ČSN ISO 690 a v Slovenskej republike zase STN ISO 690.

VI. 1. Citácie podľa normy ISO 690

VI. 1.1. ISO 690 Dokumentácia - Bibliografické odkazy

Nasledovať bude príklad, aké konkrétne informácie a v akom poradí sa majú nachádzať pri citovaní použitím normy ISO 690.

Poradie pri citovaní, príklad je hneď vedľa prvku:

- Autor - Tim Falla, Paul A. Davies
- Názov - Solutions Pre-Intermediate Student’s Book
- Poradie vydania - Druhé vydanie
- Miesto vydania, vydavateľ - Oxford University Press
- Dátum vydania - 2012
- Rozsah - 136 s.
- Štandardné číslo ISBN/ISSN - 9780194552875

Výsledný bibliografický odkaz:

Tim Falla, Paul A. Davies. *Solutions Pre-Intermediate Student's Book.* Druhé vydanie. s.l. : Oxford University Press, 2012. s. 136. 9780194552875.

VI. 1.2. ISO 690-2 Informácie a dokumentácia

Jedná sa o rozšírenie normy ISO 690 a uvádza, akým spôsobom by sa mali vytvárať bibliografické odkazy, čo sa elektronických dokumentov týka. (19) Prvky a príklad bibliografického odkazu:

- Názov webovej stránky - Česko-Slovenská filmová databáze
- URL - www.csfd.cz
- Dátum prístupu - 11. 4. 2014

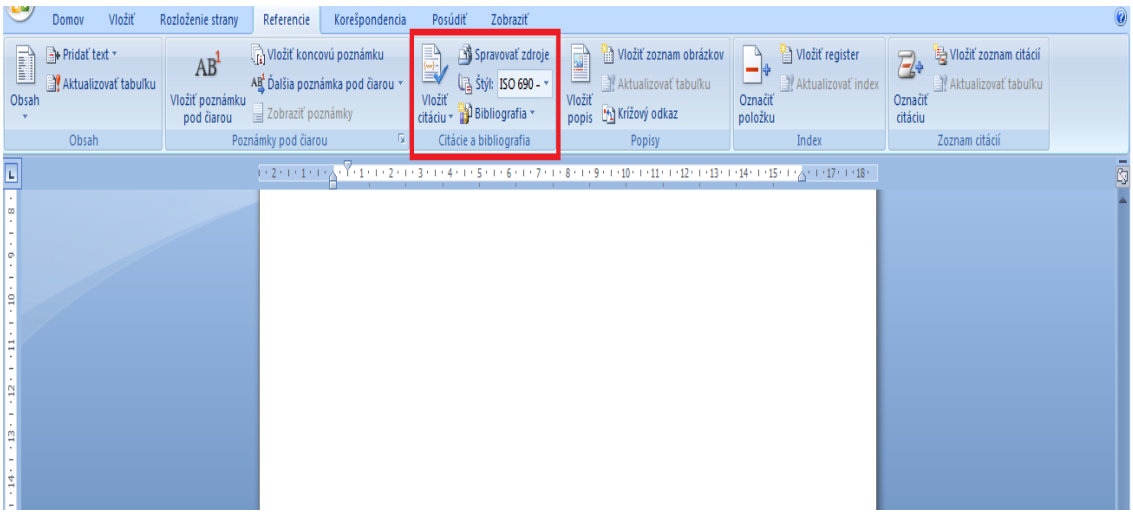
Výsledný bibliografický odkaz:

Česko-Slovenská filmová databáze. [Online] [Citace: 11. 4 2014.] www.csfd.cz.

VI. 2. Generátor citácií

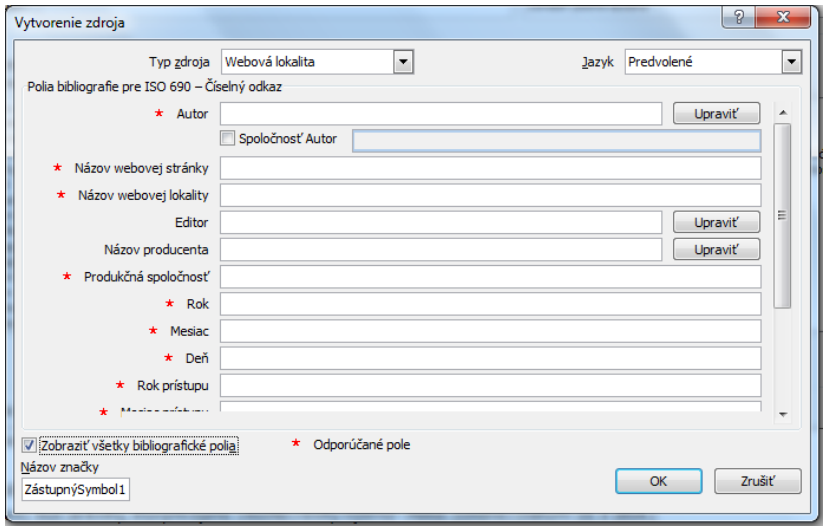
V dnešnej dobe, kedy sú informačné technológie na oveľa pokročilejšej úrovni, ako v minulosti, existuje viacero spôsobov, ako vytvárať ľahko a rýchlo citácie do prác.

Jedným z generátorov, alebo skôr funkcií či možností, ktoré pozná a používa skoro každý je MS Office Word. Stačí si iba kliknúť na kartu referencie, vybrať štýl citovania a následne vytvoriť nový zdroj. Tento ako keby návod je zachytený na obrázku č. 1.



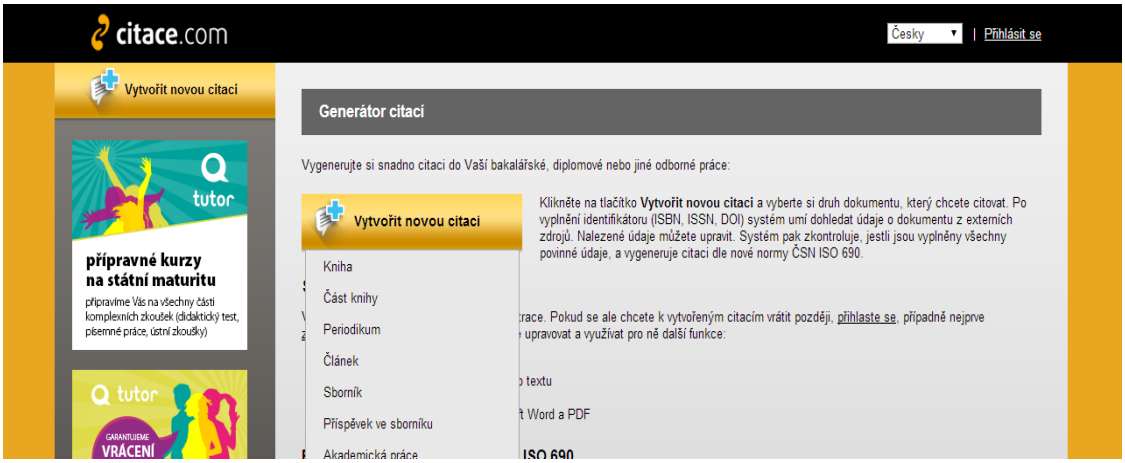
Obrázok 1: Karta referencie

Následne sa otvorí dialógové okno, v ktorom stačí vyplniť požadované polia a potom už iba vložiť bibliografiu, ktorá bude obsahovať vygenerované bibliografické odkazy, podľa vybraného štýlu. Znázornené na obrázku č. 2.



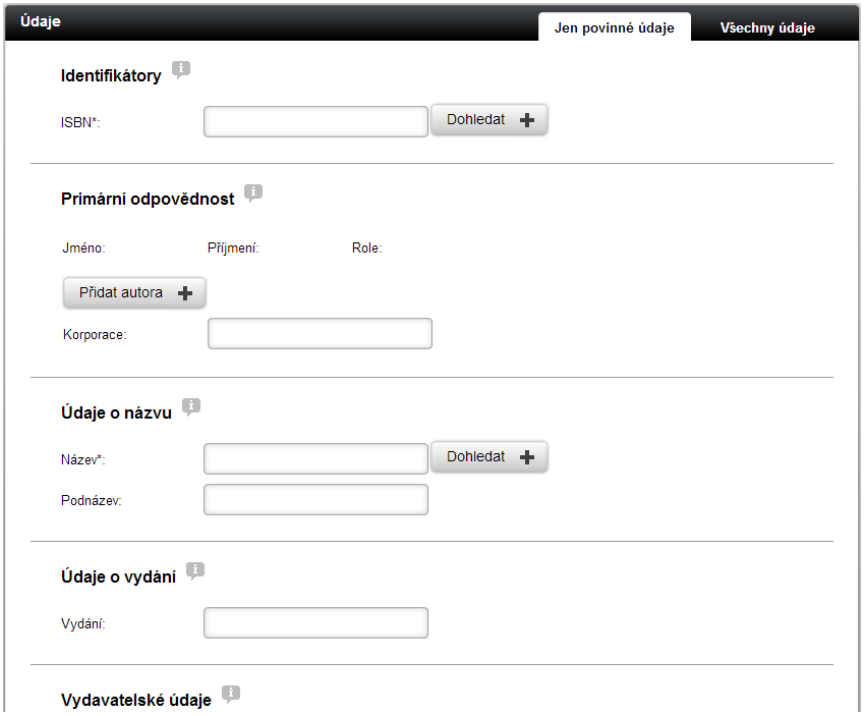
Obrázok 2: Vytvorenie nového zdroja

Druhý známy generátor sa nachádza na stránke citace.com. (20) Funguje podobne ako vytváranie citácií cez Word. Na začiatku si vyberiete druh dokumentu knihu, patent, elektronickú knihu alebo iné. Viď obrázok č. 3.



Obrázok 3: Vytvorenie citáte cez citace.com

A ďalej vás to presmeruje na stránku kde sa vypĺňajú už konkrétne údaje, ako v dialógovom okne pri vytváraní nového zdroja vo Worde. Čo ukazuje obrázok č. 4. Klasicky je tam pole pre autora, názov, ISBN či údaje o vydaní.



Obrázok 4: Údaje pre novú citáciu

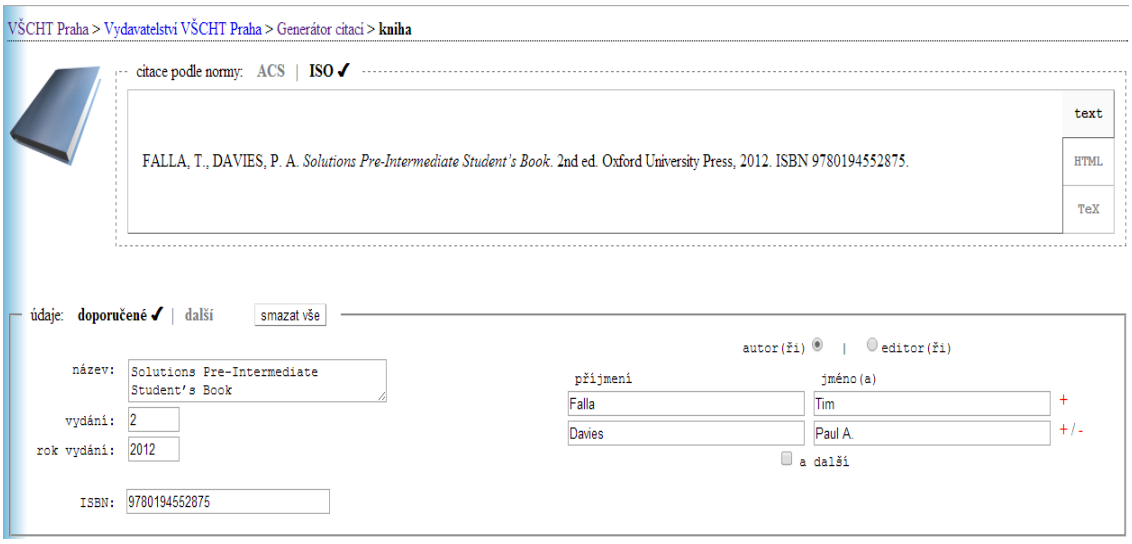
Avšak podoba je tam určite veľká, ale taktiež sa nájdu aj nejaké tie rozdiely. A to napríklad štýl citovania. Kde vo Worde je na výber z viacerých štýlov, tak na druhú stranu tento generátor používa iba ISO 690 a ISO 690-2.

Ďalším ale určite nie posledným je generátor citácií Vysokej školy chemicko-technologickej v Prahe. Tento generátor vydavateľstva VŠCHT Praha je verejnosti prístupný na http://vydavatelstvi.vscht.cz/apps/uid_ea-002/ odkiaľ budú čerpané aj všetky nasledujúce údaje. (21) Po kliknutí na odkaz je hneď jasné, že jediná funkcia webu je vytváranie citácií. Ako aj v predchádzajúcich prípadoch tak aj tu je na výber napríklad citovanie knihy, zborníku, článku a mnohých iných. Na spodnej časti stránky sa nachádza aj jeden užitočný tip: ,,U norem jsou povinné údaje opravdu povinné, u patentů podle úrovně jejich přiznání (tj. např. u mezinárodních se neuvádí země). U ostatních zdrojů jsou doporučené údaje závazné, pokud jsou k dispozici (převážně data z tiráže knihy), jinak se nemusí uvádět (mimo jiné není nutné uvádět pořadí vydání, pokud je první, jinak ano). V extrémních případech jsou jedinými skutečně povinnými údaji název citovaného dokumentu a přibližná doba jeho uveřejnění (ta se pak uvádí v hranatých závorkách ve tvaru [1971 nebo 1972], [ca1960], [197-], [18-] a podobně). U elektronických zdrojů podléhajících změnám (typicky některé online dokumenty), případně u těch, u nichž se nepodaří zjistit datum uveřejnění, se závazně uvádí datum citace“. Celá situácia je zachytená na obrázku č. 5.



Obrázok 5: Úvodná strana generátoru citácií

Po výbere typu, napríklad kniha, sa postup skoro vôbec nelíši od postupu vo Worde. Nasleduje presmerovanie už na generátor, ktorý pripomína dialógové okno nového zdroja vo Worde. Tak isto je potrebné vyplniť autora, názov, miesto vydania atď. Ako môžeme vidieť na obrázku č. 6 nižšie, tak je na výber z dvoch citačných noriem, ACS a ISO. A okrem iného je možnosť výberu formátu citácie medzi obyčajným textom, HTML a teX.



Obrázok 6: Generátor citácií VŠCHT Praha

VII. Nástroje na odhaľovanie plagiátorstva

V dnešnej, takpovediac pokročilej informačnej dobe, autori už skoro vôbec nie sú limitovaný a kedykoľvek a kdekoľvek môžu získať prístup k rôznym zdrojom. Toto všetko je možné najmä kvôli internetu, jeho rýchlemu vývinu a samozrejme aj preto, že s dostupnosťou zariadenia , ktoré je schopné pripojenia na internet, nie je žiaden problém. Či sa jedná o notebook, počítač, tablet alebo iné. Z internetu sa pomaly, ale isto stáva najlepší a nevyčerpateľný zdroj aký existuje. Vyhľadávanie na internete je pomerne veľmi jednoduché, jednak kvôli vyhľadávačom, ale aj množstvu informácií, ktoré sú verejnosti dostupné. K dispozícii sú rôzne knihy v elektronických formách, články, publikácie, či mnohé ďalšie iné. Preto sa aj ten najväčší neznalec, nebude mať vôbec žiaden problém naučiť, ako správne a efektívne vyhľadávať.

Toto všetko dáva za vzniku aj tomu, že autori sa veľmi ľahko uchýlia k plagiátorstvu. A keďže prác, ktorých treba skontrolovať, či nie sú skopírované je čím ďalej tým viacej, tak také skontrolovanie nemôže mať na svedomí len človek. Za potreby je tu aj nástroj, ktorý práce porovnáva s inými prácami, či internetovými zdrojmi. Nástrojov, ktoré sú určené na detekciu plagiátorstva je mnoho. Či už platené alebo zdarma. Príklady nástrojov na odhaľovanie plagiátorstva, na webe infogram.cz: (22)

- Turnitin
- Copy Catch
- Essay Verification engine (EVE2)
- WCopyfind
- Scan My Essay
- Theses

Nasleduje krátky popis niektorých vybraných detekčných nástrojov.

VII. 1. Turnitin

Turnitin je svetovo známa organizácia, ktorá sa zaoberá hodnotením a zlepšovaním vzdelávania študentov. Poskytuje overovanie originality a spätnú väzbu hodnoteným študentom. Jedná sa o jednu z najviac rozšírených aplikácií na svete v tomto smere. Turnitin je používaný viac ako 10 000 inštitúciami v 135 krajinách sveta. Jednou z mnohých poskytovaných služieb je iThenticate, pre detekciu plagiátov. (23)

VII. 1.1. iThenticate

Je software, ktorý slúži na vyhľadávanie plagiátov. Vydavateľom, spoločnostiam a firmám umožňuje okamžitú možnosť, aby mohli overiť pôvod či originalitu nie len dokumentov ale aj rukopisov. A taktiež umožňuje zistiť aj to, či dokumenty firiem alebo spoločností ktoré sú ich duševným vlastníctvom, nie sú nejakým spôsobom neoprávnene použité v internetovom prostredí. (22) Cenník pre overenie dokumentov: (24)

- 50$ - získanie 1 kreditu, 1 kredit je účtovaný za odovzdanie hocijakého rukopisu do 25 000 slov.
- 150$ - získanie 3 kreditov, možnosť odovzdať 1 - 3 rukopisy do 75 000 slov.
- 237,5$ - získanie 5 kreditov, možnosť odovzdať 1 - 5 rukopisov do 125 000 slov.
- 450$ - získanie 10 kreditov, možnosť odovzdať 1 -10 rukopisov do 250 000 slov.

VII. 2. Copy catch

Copy Catch je systém ktorý taktiež slúži na detekciu plagiátov, najmä v elektronických materiáloch. Porovnáva vybrané dokumenty (napríklad odovzdané prace v jednom roku, alebo všetky odovzdané práce) a následne zobrazuje stupeň podobnosti medzi nimi. (22)

VII. 3. Essay verification engine

Essay verification engine alebo aj EVE2 je nástroj, ktorý umožňuje rôznym profesorom a učiteľom na viacerých úrovniach vzdelávacieho systému na pomoc v rozhodovaní, či sa ich študenti stali plagiátormi kopírovaním, alebo čiastočným prevzatím materiálov z webu. EVE2 pri svojej kontrole podporuje formáty, ako obyčajný text, MS Office Word a Corel Word Perfect. Výstupom je vlastne odkaz na webovú stránku, z ktorej mohol študent potenciálne plagiovať. Nástroj EVE2 bol dostatočne vyvinutý na to, aby výstupné odkazy zodpovedali skutočne tým stránkam, z ktorých mohol študent kopírovať.

V programe prebieha veľké množstvo zložitých vyhľadávaní, aby bol schopný nájsť zdroj z akejkoľvek internetovej stránky. Dalo by sa konštatovať, že v dnešnej dobe by mohlo byť technicky nemožné a s dnešnými technológiami a prístupmi naprogramovať

program, ktorý by bol schopný prehľadať každý web na celom internete. EVE2 sa však priblížil k tomu zakomponovaním pokročilých vyhľadávacích nástrojov schopných identifikovať podozrivé stránky.

Akonáhle je vyhľadávanie ukončené, tak je učiteľovi poskytnutá úplná správa na každý list, ktorý obsahoval známky plagiátorstva, zahŕňajúc percentuálnu časť práce, ktorá bola splagiovaná. Výseky, ktoré vykazujú známky plagiátorstva sú zvýraznené červenou farbou. Cena programu je 29,99$. (25)

VII. 4. WCopyfind

WCopyfind je program, ktorý slúži na detekciu plagiátorstva v obsahu dokumentov. Porovnáva viacero dokumentov a hľadá medzi nimi podobnosti. Vyberá časti z týchto dokumentov a skúma podobnosť jednotlivých použitých fráz. Ak WCopyfind nájde zhody v takýchto vybraných dokumentoch, tak následne generuje HTML správu. Táto správa potom obsahuje text dokumentu, s podčiarknutými frázami ktoré sa zhodujú.

WCopyfind podporuje nasledujúce formáty, v ktorých je schopný vyhľadávať a porovnávať ich: .docx, .doc, .txt, .htm, .html a .pdf. Použitie iných formátov je technicky možné, ale program pri nich nezaručuje správnu a presnú funkčnosť. Na druhú stranu, program nie je schopný prehľadávať weby na internete, aby našiel zhodné dokumenty. Programu je za potreby označiť konkrétne dokumenty, ktoré má porovnávať. (26)

WCopyfind je voľne stiahnuteľný pre každého bez poplatkov. Licencovaný je pod GNU Public License, čo vlastne znamená to, že s ním môžete robiť čokoľvek uznáte za vhodné. Jedinou výnimkou je pokus o predaj niekomu inému, ten je samozrejme zakázaný. (27)

VII. 5. Scan my Essay

Web Scan my Essay poskytuje na svojich stránkach program s názvom Viper. Jedná sa o program, ktorý poskytuje kontrolu čo sa plagiátorstva týka. Porovnáva do akej miery sa texty zhodujú a k dispozícii je taktiež zdarma. Na oficiálnej stránke sa nachádza tabuľka, v ktorej je Viper porovnaný so systémom WriteCheck od Turnitinu. Viď tabuľka č. 1. (28)

Fakt	Viper	WriteCheck
Cez 10 miliárd zoskenovaných zdrojov	Áno	Áno
Neobmedzená dĺžka dokumentu	Áno	Nie
Neobmedzené opätovné odovzdávanie	Áno	Nie
Porovnávanie dokumentov bok po boku	Áno	Nie
Odkazy na plagiovanú prácu	Áno	Nie
Bez poplatkov	Áno	nie

Tabuľka 1: Porovnanie Viper a WriteCheck

VII. 6. Theses

Theses.cz je taktiež nástroj, ktorý slúži na detekciu plagiátov v záverečných prácach. Vyvíjaný a prevádzkovaný je Masarykovou univerzitou. Používaný je vysokými školami a univerzitami (nie len v Českej republike) a slúži im aj ako národný register záverečných prác. Poskytuje rôzne informácie o týchto prácach, ako napríklad názov či autora. Systém taktiež povoľuje zapojeným školám vkladať práce a následne aj vyhľadávať medzi nimi potenciálne plagiáty. (29)

Škôl ktoré používajú Theses je celkovo 41 a patria medzi ne napríklad nasledujúce:

- Vysoká škola ekonomická v Praze
- Masarykova univerzita
- Ekonomická univerzita v Bratislave
- London International Graduate School
- Univerzita Karlova v Praze

Tím z Masarykovej univerzity získal za tento projekt v roku 2009 cenu Inforum, čo tiež poukazuje na to, že sa jedná o kvalitný detekčný nástroj. Celý projekt vznikol na základe dopytu vysokých škôl po nástroji, ktorý bude úspešne odhaľovať plagiátorov záverečných prác. Do systému bolo zatiaľ pridaných skoro 73 000 záverečných prác. Práce sú porovnávané v databáze, ktorá obsahuje 1,3 milióna dokumentov.

Informatici z Masarykovej univerzity vytvorili aj ďalší systém podobný Thesesu s názvom Odevzdej.cz, ktorý bol spustený v roku 2009. Od Thesesu sa líši tým, že

umožňuje vyhľadávať plagiáty nie medzi záverečnými prácami, ale aj medzi seminárnymi prácami, referátmi, slohmi a inými. (30)

Inštitúcií, ktoré používajú systém Odevzdej.cz je spolu 38 a patria medzi ne napríklad:

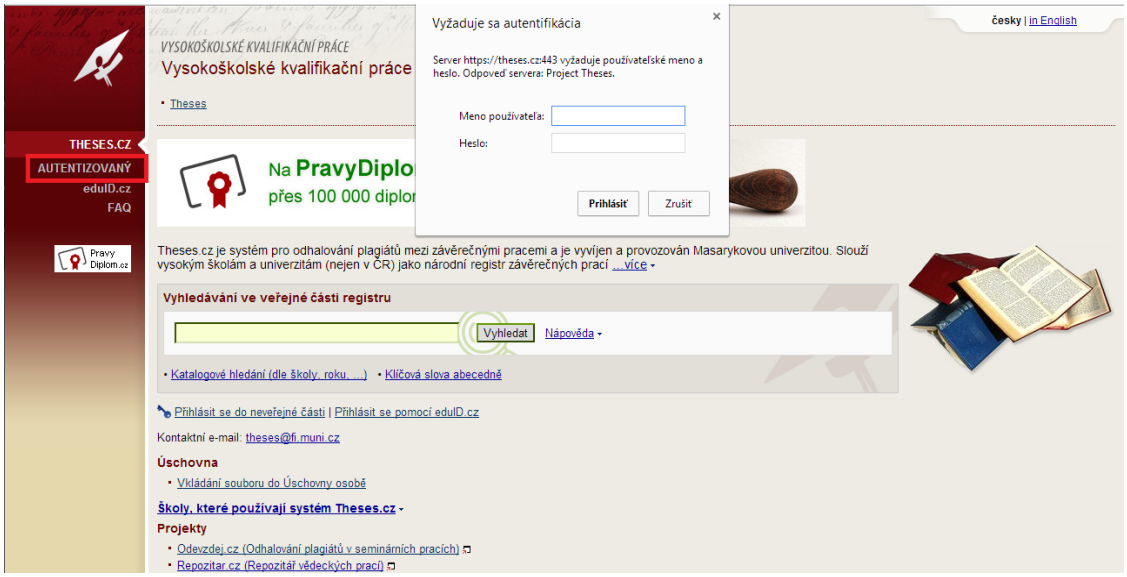
- Univerzita Hradec Králové
- Škoda Auto, a. s. Vysoká škola
- Gymnázium Brno - Řečkovice
- Ministerstvo práce, sociálnych vecí a rodiny Slovenskej republiky

VII. 6.1. Praktické použitie Thesesu

V nasledujúcej podkapitole bude vysvetlené, ako vkladať do systému Theses práce, ktoré majú byť overené na plagiátorstvo. Po ukážke ako spraviť tento základný úkon, bude do systému vložený dokument s textom približne na trištvrte strany. Tento dokument sa v tomto malom pokuse bude tváriť ako originál, s ktorým budú porovnávané jeho plagiátorské verzie. Zdrojom textu v dokumentu je wikipédia. Wikipédia síce nie je dôveryhodný zdroj, ale v tomto prípade ide len o to, aké zhody či nezhody nájde systém v texte. Čiže kvalita textu nehrá až takú veľkú rolu. Všetky obrázky, ktoré budú nasledovať a taktiež aj operácie na nich pochádzajú zo stránky www.theses.cz.

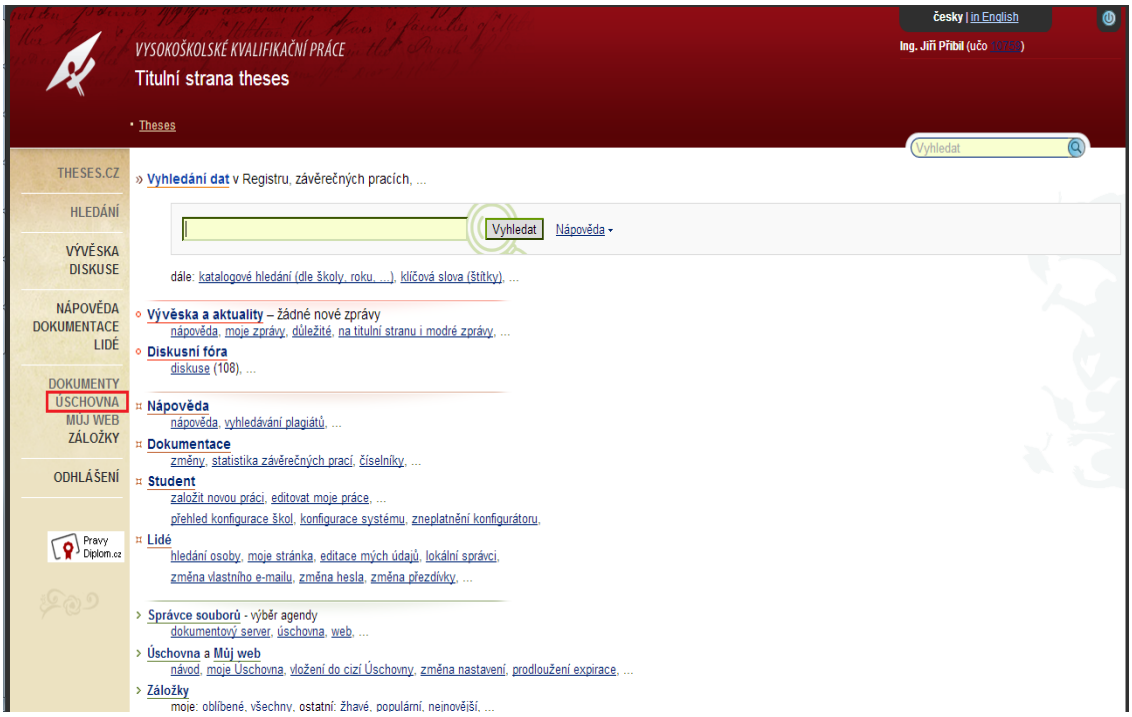
Poskytnutý mi bol užívateľský účet, v rámci ktorého budú prebiehať ukážky použitia a príklad overovania.

Do systému je možné sa prihlásiť na domovskej stránke Thesesu kliknutím na „Autentizovaný“. Po kliknutí vyskočí dialógové okno v ktorom je potrebné vyplniť meno používateľa a heslo pre ďalší prístup.



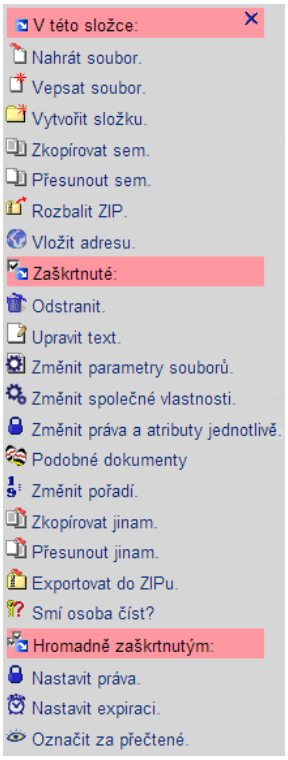
Obrázok 7: Prihlasovanie do systému

Po prihlásení sa objaví úvodná strana. Ďalším krokom je kliknutie na „Úschovna“ na ľavom panely.



Obrázok 8: Titulná strana

Úschovňa ponúka viacero akcií, aké je možné robiť s dokumentmi v systéme. Ako napríklad: nahrávanie súboru, kopírovanie, upravovanie textu, porovnávanie a ďalšie, ktoré sú zobrazené na obrázku č. 9.

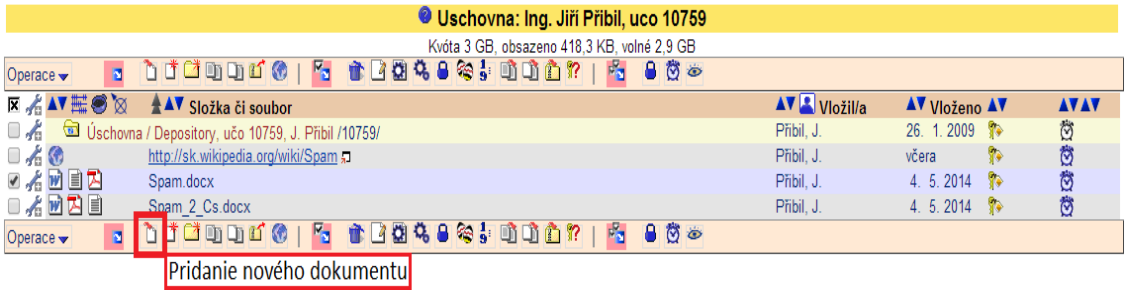


Obrázok 9: Možnosti úschovne

Ponuka je naozaj široká, ale pre môj pokus porovnávania budem používať iba nahrávanie nových súborov a funkciu podobné dokumenty.

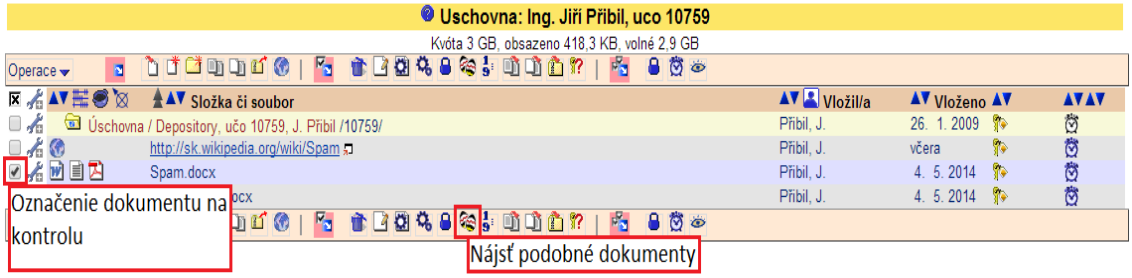
Na úplnom začiatku som do systému nahral dokument s názvom „Spam“ (príloha č. 1), ktorého obsah pochádza iba obyčajným doslovným skopírovaním časti originálu, aby som vyskúšal, či sa vyskytnú nejaké technické problémy. Ďalej som musel vložiť aj webovú stránku, z ktorej text pochádza, aby ho systém mohol porovnať nie len s nájdenými zhodnými dokumentmi ale rovno aj s originálom. Dokumenty je možné vkladať jednoducho v docx. formáte, netreba ich konvertovať ani meniť, pretože si ich systém sám zmení či už na .pdf alebo .txt (obyčajný text).

Prvým krokom v nasledujúcich pokusoch bolo nahratie súboru pomocou funkcie „Nahrát soubor z počítače“.



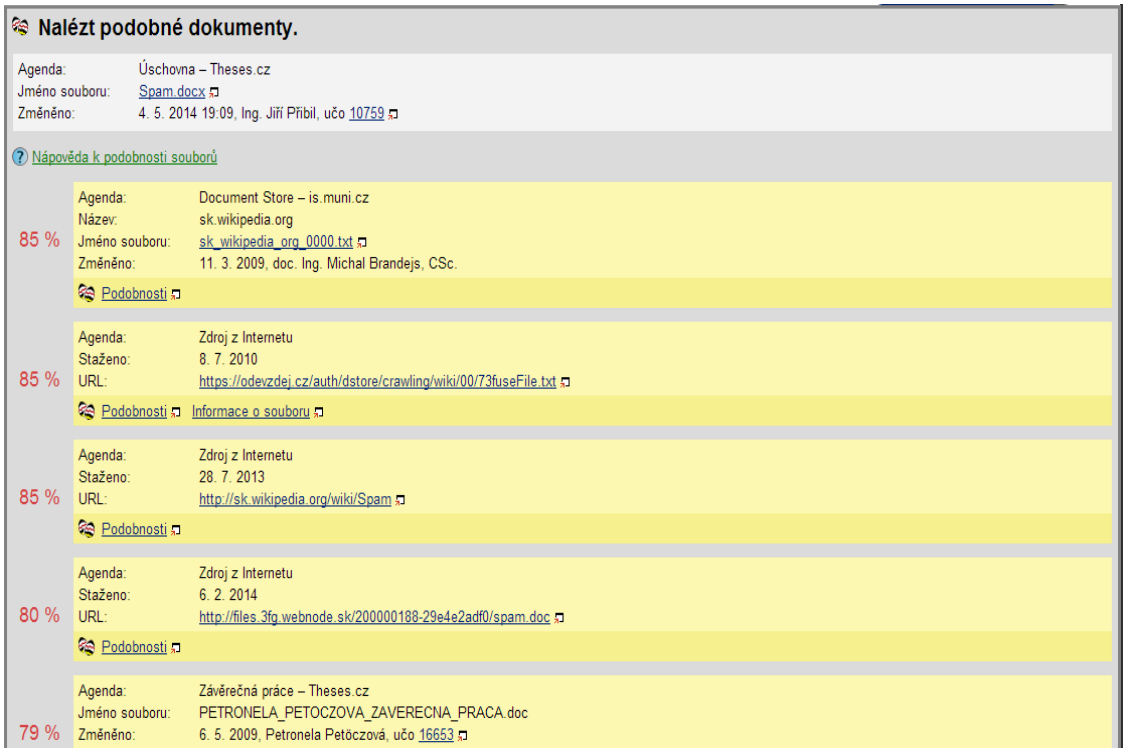
Obrázok 10: Úschovna - pridanie nového dokumentu

Druhým krokom bolo označenie požadovaného súboru ku kontrole a následné použitie funkcie „Nalézt podobné dokumenty“.



Obrázok 11: Úschovna - Označenie a hľadanie podobností

Výsledkom, okrem nájdenia zhody s originálom, bolo nájdenie internetových zdrojov alebo aj prác, v ktorých sa daný text nejakým spôsobom nachádza. Na obrázku č. 17 je možné vidieť rôzne percentuálne zhody, od 85% zhody až po nižšiu 10% zhodu.



Obrázok 12: Nájdené podobné dokumenty

Za celkom zvláštny jav sa dá považovať to, že aj napriek tomu, že som daný text v dokumente iba skopíroval, tak aj tak systém nenašiel 100% zhodu. Mojim zdrojom bola v tomto prípade wikipédia, konkrétne http://sk.wikipedia.org/wiki/Spam. V zozname vyššie je to zhora, tretia nájdená podobnosť a iba 85%. Pri všetkých nájdených dokumentoch je možné si kliknúť na „Podobnosti“, kde si môžeme pozrieť, ktoré časti textu systém našiel v nájdených dokumentoch. Na obrázku č. 13 nižšie je možné vidieť že celý môj text až na jeden odstavec systém označil červenou farbou, ktorá znamená zhodu. Avšak ma zaujalo to, prečo práve ten kúsok textu označený nie je, keď je tak isto ako zvyšok textu jednoducho skopírovaný cez ctrl+c a ctrl+v. Ako jediné vysvetlenie ma napadlo, že neoznačený úsek v texte vlastne nie je v samostatnom odstavci (pričom na stránke wikipédia v samostatnom odstavci je) čo možno oklamalo systém a tým pádom ho neoznačil za skopírovaný.

Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

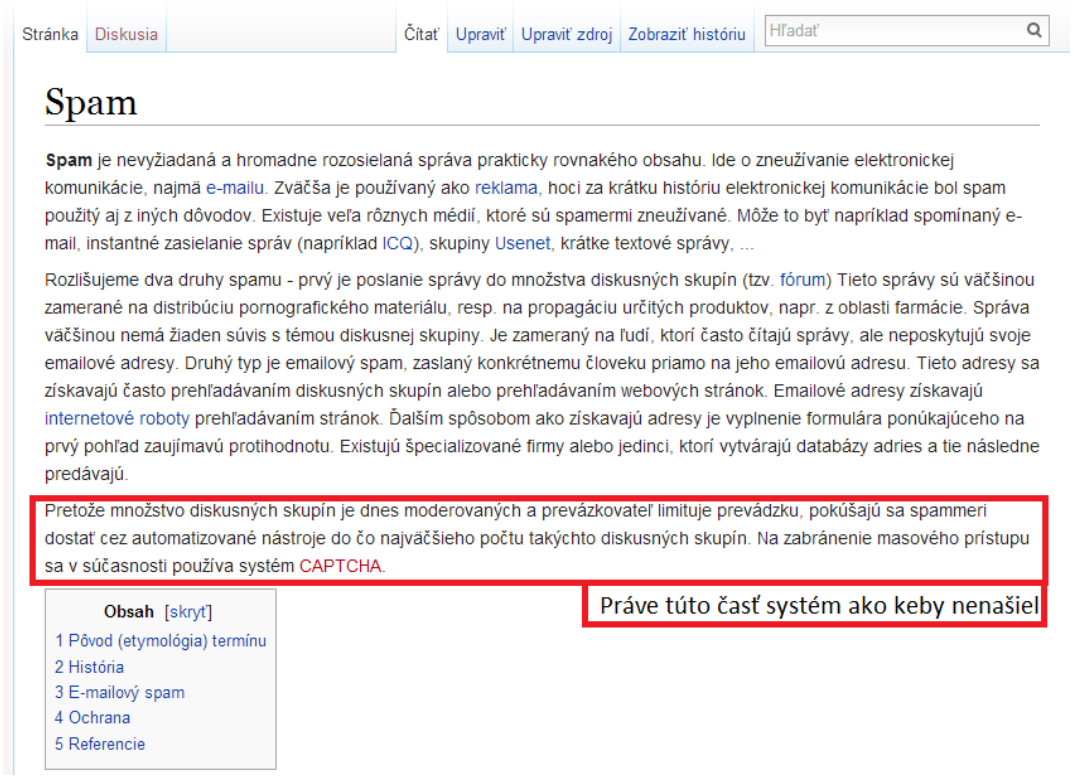
Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré sú spamermi zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy, ...

Rozlišujeme dva druhy spamu - prvý je poslanie správy do množstva diskusných skupín (tzv. fórum) Tieto správy sú väčšinou zamerané na distribúciu pornografického materiálu, resp. na propagáciu určitých produktov, napr. z oblasti farmácie. Správa väčšinou nemá žiaden súvis s témou diskusnej skupiny. Je zameraný na ľudí, ktorí často čítajú správy, ale neposkytujú svoje emailové adresy. Druhý typ je emailový spam, zaslaný konkrétnemu človeku priamo na jeho emailovú adresu. Tieto adresy sa získavajú často prehľadávaním diskusných skupín alebo prehľadávaním webových stránok. Emailové adresy získavajú internetové roboty prehľadávaním stránok. Ďalším spôsobom ako získavajú adresy je vyplnenie formulára ponúkajúceho na prvý pohľad zaujímavú protihodnotu. Existujú špecializované firmy alebo jedinci, ktorí vytvárajú databázy adries a tie následne predávajú.

Pretože množstvo diskusných skupín je dnes moderovaných a prevádzkovateľ limituje prevádzku, pokúšajú sa spameri dostať cez automatizované nástroje do čo najväčšieho počtu takýchto diskusných skupín. Na zabránenie masového prístupu sa v súčasnosti používa systém CAPTCHA.
Ochrana proti spamu
je komplikovaná. Problém je v tom, že neexistuje ostrá hranica medzi spamom a užitočnými emailami a neexistujú príznaky, ktoré by jednoznačne indikovali, že ide o spam. Následkom toho nie je možné vytvoriť prostriedok, ktorý by na sto percent likvidoval spam a na sto percent prepúšťal ostatné emaily.

Obrázok 13: Podobnosti textu

Na ďalšom obrázku je zobrazený text z wikipédia na ktorom jasne vidieť, že neoznačená časť textu sa tam predsa nachádza a teda, podľa správnosti by táto časť mala byť taktiež označená červeným.



Obrázok 14: Text z wikipédie

Táto skutočnosť mi nedala, aby som s touto verziou plagiátu neskúsil ešte jeden pokus. V dokumente som podozrivú časť textu dal do samostatného odstavca a skúsil či ju systém označí za podobnú. Avšak výsledok bol stále taký istý, podozrivá časť textu nebola ani v tomto prípade označená za podobnú či skopírovanú. Odstavec je označený za plagiát len v tom prípade, keď sa dokument porovná s dokumentom v ktorom odstavec nie je ešte vytvorený, čiže ak sa porovná s prílohou č. 1.

Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré sú spamermi zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy, ...

Rozlišujeme dva druhy spamu - prvý je poslanie správy do množstva diskusných skupín (tzv. fórum) Tieto správy sú väčšinou zamerané na distribúciu pornografického materiálu, resp. na propagáciu určitých produktov, napr. z oblasti farmácie. Správa väčšinou nemá žiaden súvis s témou diskusnej skupiny. Je zameraný na ľudí, ktorí často čítajú správy, ale neposkytujú svoje emailové adresy. Druhý typ je emailový spam, zaslaný konkrétnemu človeku priamo na jeho emailovú adresu. Tieto adresy sa získavajú často prehľadávaním diskusných skupín alebo prehľadávaním webových stránok. Emailové adresy získavajú internetové roboty prehľadávaním stránok. Ďalším spôsobom ako získavajú adresy je vyplnenie formulára ponúkajúceho na prvý pohľad zaujímavú protihodnotu. Existujú špecializované firmy alebo jedinci, ktorí vytvárajú databázy adries a tie následne predávajú.

Pretože množstvo diskusných skupín je dnes moderovaných a prevádzkovateľ limituje prevádzku, pokúšajú sa spameri dostať cez automatizované nástroje do čo najväčšieho počtu takýchto diskusných skupín. Na zabránenie masového prístupu sa v súčasnosti používa systém CAPTCHA.

Ochrana proti spamu je komplikovaná. Problém je v tom, že neexistuje ostrá hranica medzi spamom a užitočnými emailami a neexistujú príznaky, ktoré by jednoznačne indikovali, že ide o spam. Následkom toho nie je možné vytvoriť prostriedok, ktorý by na sto percent likvidoval spam a na sto percent prepúšťal ostatné emaily.

Obrázok 15: Podobnosti textu po úprave

Pre ďalší pokus či skúšku fungovania systému Theses som dokument „Spam“ (príloha č. 1) poskytol ďalším ľuďom aby vytvorili plagiát. Avšak vo všetkých prípadoch bolo nutné aby som do nových plagiátov pridal celý skopírovaný odstavec, pretože ich bez toho systém nenašiel ako zhodné dokumenty. Bez skopírovanej časti bolo percento zhody zrejme zanedbateľné a systém vypísal nasledujúcu správu: „ K vloženému souboru nebyl v databázi nalezen žádný podobný dokument“. Nasledujúce dokumenty boli porovnávané s dokumentom „Spam“ (príloha č. 1), nie so stránkou wikipédia.

V druhom prípade (príloha č. 2) boli použité nasledujúce spôsoby:

- Zmena štýlu

Slová boli zamieňané za ich synonymá a celý text bol trošku preštylizovaný, ale s tým, že sa zachovala pôvodná myšlienka.

- Zmena slovosledu

Vety nezostali po použití synoným vo svojom pôvodnom tvare ale boli pozmenené aj pozície jednotlivých slov.

Ako môžeme vidieť na obrázku č. 16, tak nájdených podobností je naozaj len minimum a ako som už spomínal tak na to, aby systém vôbec našiel zhodný dokument a tým pádom aby bol aj možný náhľad na podobnosti, bolo za potreby do kontrolovaného dokumentu vložiť doslovným skopírovaním odstavec z originálu. Skopírovaný odstavec je celý označený červenou farbou a netreba ho vôbec brať do úvahy, keďže bol pridaný len za tým účelom, aby bolo možné nahliadnuť na podobnosti. Ostatných pár slov, ktoré sú označené, sú také slová, ktoré sa v tomto poradí aj znení nachádzali aj v pôvodnom dokumente. Čiže sa dá povedať, že označené slová či časti viet boli ponechané na svojom pôvodnom mieste a neboli žiadnym spôsobom zmenené. Zvyšok textu ktorý bol buď zmenený cez synonymá, slovosled alebo aj ich kombináciou, označený nie je.

Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré sú spamermi zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy,
...

Spam je formou nevyžiadanej a hromadne rozosielanej správy s takmer rovnakým jadrom. Jedná sa hlavne o zneužitie formy el. komunikácie, zväčša mailu. Najčastejšie je používaný formou reklamy, ale počas nedlhej histórie el. komunikácie bol používaný aj pre iné účely. Je veľa rozličných druhov médií, ktoré spameri využívajú. Najčastejšie je to mail, rôzne chaty (ako MSN messenger), sms-ky. A skupiny usenet.

Poznáme 2 druhy spamu - tým prvým je posielanie spamu do rôznych fór. Takéto rozosielanie je zväčša zacielené na šírenie pornografického materiálu, určitých produktov, napr. z oblasti lekárnictva. Správa nijak nesúvisí s témou skupiny diskusných fór. Zameriava
sa na ľudí ktorí často čítajú
správy,
ale svoje emaily neposkytujú. Druhým typom je mailový spam, ktorý sa posiela konkrétnej osobe na mailovú adresu. Tieto adresy sa dajú získať prehľadávaním diskusných skupín a webových stránok. Hľadajú ich taktiež internetové roboty ktoré prehľadávajú stránky. Ďalším zo spôsobov je ponúknutie osobe vyplniť formulár za zaujímavú protihodnotu. Taktiež jestvujú špeciálne firmy alebo jednotlivci, ktoré vytvárajú databázy mailových adries a potom ich predávajú.

Keďže v súčasnosti je väčšina diskusných skupín moderovaných, a ich prevádzkovateľ limituje prevádzku, spameri sa snažia
dostať cez automatizované nástroje do čo najväčšieho počtu
diskusných
fór. Systém CAPTCHA bol vyvinutý práve na zabránenie takýchto útokov. Ochrana proti spamu je dosť zložitá. Problém nastáva v tom, že nevieme presne určiť konkrétnu hranicu medzi spamom a užitočnými emailmi, a nie sú žiadne ukazovatele ktoré by jasne stanovil, že sa jedná o spam. V dôsledku toho je nemožné vytvoriť nástroj, ktorý by na 100% likvidoval spam, a ostatné emaily púšťal.

Obrázok 16: Nájdené podobnosti v dokumente č. 2

V ďalšom čiže treťom prípade (príloha č. 3) boli použité nasledujúce spôsoby vytvárania plagiátu:

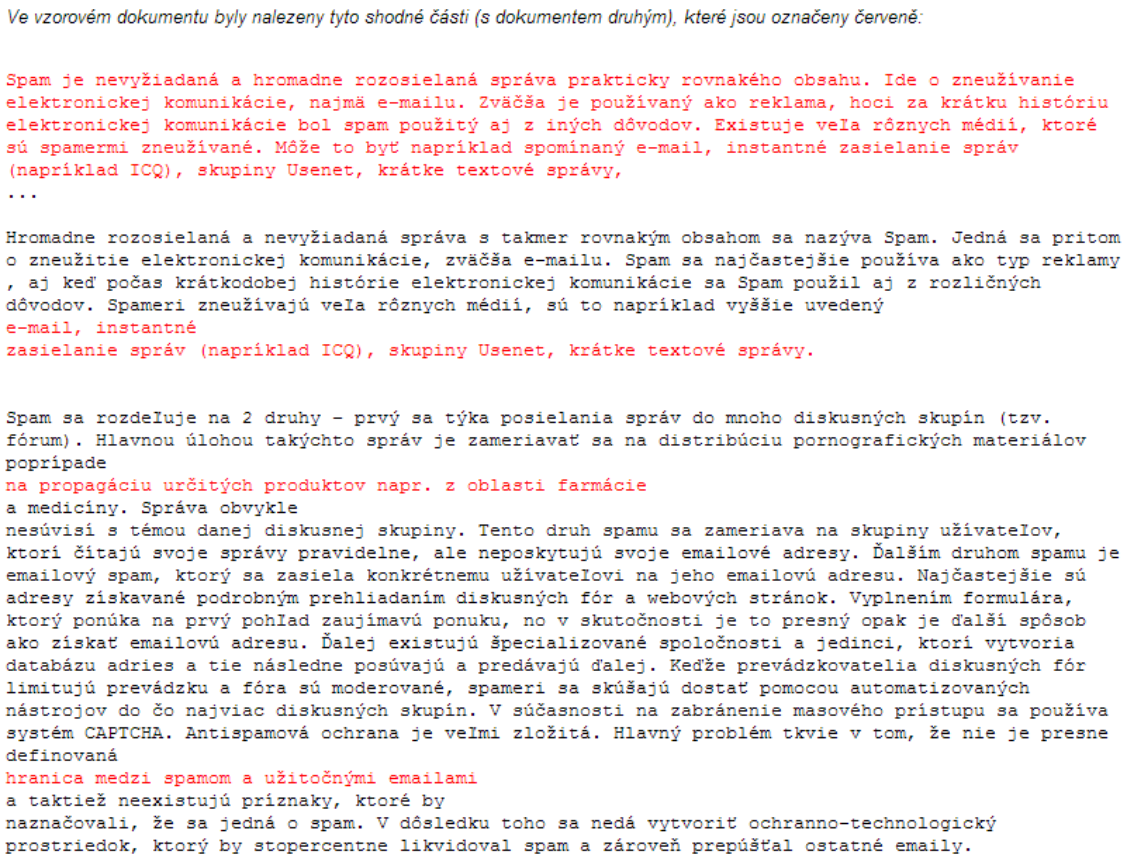
- Zmena slovosledu

Taktiež zmenené poradie slov ako v predchádzajúcom prípade.

- Zmena štýlu

Niektoré slová boli nahradené ich synonymami ale myšlienka zostala taktiež zachovaná.

Na obrázku č. 17 môžeme vidieť nájdené podobnosti po použití vyššie uvedených spôsobov. Prvý odstavec je tak isto ako v predchádzajúcom prípade použitý len za účelom, aby systém našiel zhodný dokument. Po nahliadnutí môžeme vidieť, že označené časti sa líšia od označených častí v predchádzajúcom prípade. Avšak označené sú tiež kvôli tomu istému dôvodu. A to kvôli tomu že označené časti viet sa nachádzajú v tomto istom znení a na tom istom mieste ako v pôvodnom dokumente „Spam“ (príloha č. 1).



Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré sú spamermi zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy,
...

Hromadne rozosielaná a nevyžiadaná správa s takmer rovnakým obsahom sa nazýva Spam. Jedná sa pritom o zneužitie elektronickej komunikácie, zväčša e-mailu. Spam sa najčastejšie používa ako typ reklamy , aj keď počas krátkodobej histórie elektronickej komunikácie sa Spam použil aj z rozličných dôvodov. Spameri zneužívajú veľa rôznych médií, sú to napríklad vyššie uvedený
e-mail, instantné
zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy.

Spam sa rozdeľuje na 2 druhy – prvý sa týka posielania správ do mnoho diskusných skupín (tzv. fórum). Hlavnou úlohou takýchto správ je zameriavať sa na distribúciu pornografických materiálov poprípade
na propagáciu určitých produktov napr. z oblasti farmácie
a medicíny. Správa obvykle
nesúvisí s témou danej diskusnej skupiny. Tento druh spamu sa zameriava na skupiny užívateľov, ktorí čítajú svoje správy pravidelne, ale neposkytujú svoje emailové adresy. Ďalším druhom spamu je emailový spam, ktorý sa zasiela konkrétnemu užívateľovi na jeho emailovú adresu. Najčastejšie sú adresy získavané podrobným prehliadaním diskusných fór a webových stránok. Vyplnením formulára, ktorý ponúka na prvý pohľad zaujímavú ponuku, no v skutočnosti je to presný opak je ďalší spôsob ako získať emailovú adresu. Ďalej existujú špecializované spoločnosti a jedinci, ktorí vytvoria databázu adries a tie následne posúvajú a predávajú ďalej. Keďže prevádzkovatelia diskusných fór limitujú prevádzku a fóra sú moderované, spameri sa skúšajú dostať pomocou automatizovaných nástrojov do čo najviac diskusných skupín. V súčasnosti na zabránenie masového prístupu sa používa systém CAPTCHA. Antispamová ochrana je veľmi zložitá. Hlavný problém tkvie v tom, že nie je presne definovaná
hranica medzi spamom a užitočnými emailami
a taktiež neexistujú príznaky, ktoré by
naznačovali, že sa jedná o spam. V dôsledku toho sa nedá vytvoriť ochranno-technologický prostriedok, ktorý by stopercentne likvidoval spam a zároveň prepúšťal ostatné emaily.

Obrázok 17: Nájdené podobnosti v dokumente č. 3

V oboch prípadoch systém označil za podozrivé úseky také časti viet, ktoré vznikli pomocou ctrl+c a následne ctrl+v.

Štvrtý a posledný dokument (príloha č. 4), ktorý vznikol pre tento pokus používaním plagiátorských postup by som označil ako za najzaujímavejší. Spôsob ktorý bol použitý pri vytvorení tohto dokumentu je len jeden a ním je preklad zo slovenčiny do češtiny.

Pridaných bolo len veľmi málo slov, alebo odobratých, ale aj to len kvôli tomu, aby boli vety gramaticky správne.

Ako môžeme vidieť na obrázku č. 18 nižšie tak je podobností opäť minimum, až by som povedal, že na pár výnimiek systém označil iba tie slová ktoré sa či už v slovenskom alebo českom jazyku píšu rovnako. Výnimkami sú konkrétne slová napríklad, krátké, své a používá. Je zvláštne, že tieto slová označené boli a ostatné slová označené neboli. Dalo by sa to síce odôvodniť tým, že ich berie ako slová, ktoré proste patria k slovu, ktoré po nich nasleduje. Nejaké zámená alebo spojky, ako napríklad „své emailové adresy“, v tomto prípade si myslím, že to zmysel dáva. Každopádne je zaujímavé, že aj keď sa systém Theses používa na slovenských aj českých školách, tak vlastne nie je vyladený natoľko, aby rozpoznal rozdiely v tak dvoch veľmi podobných jazykoch. Je samozrejmé, že niektoré výrazy v slovenčine sú úplne neznáme pre český jazyk a naopak. Ale niektoré slová, ktoré sa líšia len v jednom mäkčeni alebo dĺžni by poznať teoreticky mohol.

Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

Spam

Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré sú spamermi zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy,
...

Spam je nevyžádaná a hromadně rozesílaná zpráva prakticky stejného obsahu. Jedná se o zneužívání elektronické komunikace, především e-mailu. Spam je nejvíce používaný jako reklama, avšak za krátkou historii elektronické komunikace byl spam použitý i z jiných důvodů. Existuje mnoho různých médií, které jsou spamermi zneužívané. Může to být například již zmiňovaný e-mail, zasílání zpráv

(například ICQ), skupiny Usenet, krátké textové
zprávy, aj.

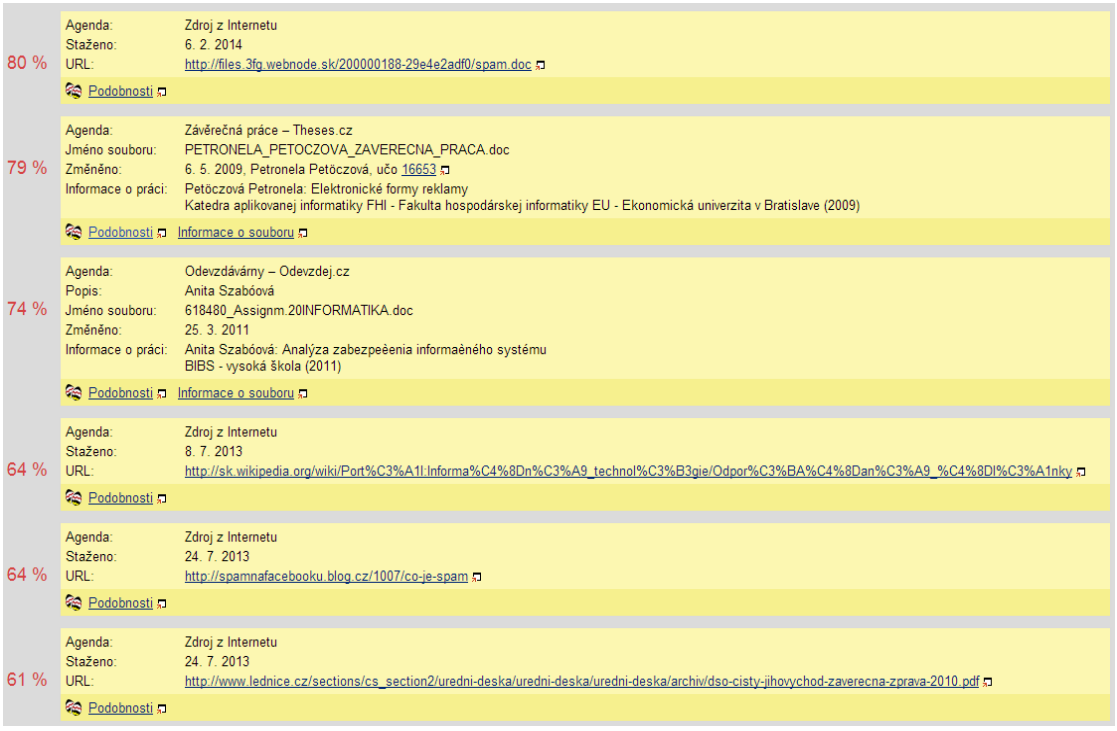
Rozlišujeme dva druhy spamu. První je posílání zprávy do většího množství diskuzních skupin (tzv. fórum). Tyto zprávy jsou většinou zaměřené na distribuci pornografického materiálu, respektive na propagaci určitých produktů, např. z oblasti farmacie. Posílaná zpráva většinou nemá žádnou souvislost s tématem diskuzní skupiny. Spam je zaměřený na lidi, kteří často čtou zprávy, ale neposkytují
své emailové adresy. Druhý typ je emailový spam, zaslaný
konkrétnímu člověku přímo na
jeho emailovou adresu. Tyto adresy se získávají hlavně prohledáváním stránek. Emailové adresy získávají internetoví roboti prohledáváním stránek. Dalším způsobem jakým získávají adresy je vyplnění formuláře nabízejícího na první pohled zajímavou protihodnotu.

Existují specializované firmy nebo jedinci, kteří vytvářejí databázi adres a ty následně prodávají. Protože množství diskuzních skupin je dnes moderovaných a provozovatel limituje provoz, pokoušejí se spameři dostat přes automatizované nástroje do co největšího počtu takovýchto diskuzních skupin. Na zabránění masového přístupu se v současnosti
používá systém CAPTCHA.

Ochrana proti spamu je komplikovaná. Problém je v tom, že neexistuje ostrá
hranice mezi spamem a
užitečnými emailami a neexistují příznaky, které by jednoznačně identifikovali, že jde o spam. Následkem toho není možné vytvořit prostředek, který by na sto procent likvidoval spam a na sto procent pouštěl ostatní emaily.

Obrázok 18: Nájdené podobnosti v dokumente č. 4

Systém Theses samozrejme našiel aj iné dokumenty v ktorých bol použitý ten istý text, ktorý som použil aj ja. Na obrázku č. 19 môžeme vidieť niektoré z takých dokumentov.



Obrázok 19: Ďalšie podobné dokumenty

Na začiatok je hneď vhodné napísať, že percentá majú kvôli tomu také vysoké hodnoty, že sa jedná o pomerne krátky dokument, ktorý nemá ani jednu stranu.

Zo zaujímavosti som si otvoril podobnosti pri práci, ktorú napísala Petöczová Petronela: Elektronické formy reklamy. Na obrázku č. 20 sa dá pekne vidieť, koľko textu bolo skopírovaného pomocou funkcie ctrl+c a ctrl+v do danej práce. Či už z wikipédie alebo z iného zdroja, kde sa tieto informácie môžu nachádzať. Skopírované časti sú opäť červenou farbou.

Ve vzorovém dokumentu byly nalezeny tyto shodné části (s dokumentem druhým), které jsou označeny červeně:

Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré

sú spamermi
zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ
(napríklad
ICQ), skupiny Usenet,
krátke textové správy, ...

Rozlišujeme dva druhy spamu - prvý je poslanie správy do množstva diskusných skupín (tzv. fórum) Tieto správy sú väčšinou zamerané na distribúciu pornografického materiálu, resp. na propagáciu určitých produktov, napr. z oblasti farmácie. Správa väčšinou nemá žiaden súvis s témou diskusnej skupiny. Je zameraný na ľudí, ktorí často čítajú správy, ale neposkytujú svoje emailové adresy. Druhý typ je emailový spam, zaslaný konkrétnemu človeku priamo na jeho emailovú
adresu.
Tieto
adresy sa získavajú často prehľadávaním diskusných skupín alebo prehľadávaním webových stránok. Emailové adresy získavajú internetové roboty prehľadávaním stránok. Ďalším spôsobom ako získavajú adresy je vyplnenie
formulára ponúkajúceho
na prvý pohľad zaujímavú protihodnotu. Existujú
špecializované firmy alebo jedinci, ktorí vytvárajú databázy adries a tie následne predávajú.

Pretože množstvo diskusných skupín je dnes moderovaných
a prevádzkovateľ limituje prevádzku,
pokúšajú sa spameri
dostať cez automatizované nástroje do čo najväčšieho počtu takýchto diskusných skupín. Na zabránenie masového prístupu sa v súčasnosti používa systém CAPTCHA.

Ochrana proti spamu je komplikovaná. Problém je v tom, že neexistuje ostrá hranica medzi spamom a užitočnými emailami a neexistujú príznaky, ktoré by jednoznačne indikovali,

že ide o spam.
Následkom toho nie je možné vytvoriť prostriedok, ktorý by na sto percent likvidoval spam a na sto percent prepúšťal ostatné emaily.

Obrázok 20: Podobnosti v záverečnej práci

Ako zhrnutie všetkých pokusov môžem povedať, že systém Theses je používaný na odhaľovanie plagiátorstva. Ale nie je všemocný, ako by si niekto mohol myslieť. Odhaľovanie doslovného kopírovania pomocou ctrl+c a ctrl+v zvláda bez problémov a takéto plagiátorstvo odhalí veľmi jednoducho a s prehľadom. Avšak stačí využiť jednoduchú kombináciu použitia synoným s následnou zmenou slovosledu a systém potom nebude schopný odhaliť časti, ktoré boli plagizované. Ako dodatok je vhodné spomenúť aj to, že vzťah slovenčiny a češtiny je systému viac menej cudzí a stačí menšia úprava slova, či už dĺžňom alebo mäkčeňom a systém ich hneď nie je schopný identifikovať ako podobné.

VIII. Záver

Hlavným cieľom mojej bakalárskej práce bolo vyskúšať prístupnosť, funkčnosť a efektivitu jedného zo systémov, ktorý slúži k odhaľovaniu plagiátorstva v záverečných prácach, či iných textových dokumentoch. Konkrétne sa jednalo o skúšku systému Theses, ktorý je prevádzkovaný týmom informatikov z Masarykovej univerzity v Brne. Celá skúška či pokus prebiehal za pomoci užívateľského účtu, ktorý mi bol poskytnutý práve pre tieto účely.

Pointa pokusu spočívala vo vytvorení dokumentu, funkciou ctrl+c a ctrl+v, ktorého obsah pochádzal z internetového zdroja. Z tohto dokumentu boli následne vytvorené ďalšie tri verzie, vlastne plagiáty, ktoré som preveril pomocou systému Theses. K vytvoreniu týchto plagiátov boli použité nasledujúce postupy:

- Zmena slovosledu

Jednotlivé slová dostali novú pozíciu vo vetách.

- Zmena štýlu

Slová vo vetách nedostali iba novú pozíciu, ale niektoré boli aj nahradené synonymami. Pripomenúť však treba, že myšlienka celého textu sa vždy zachovala, zostala nezmenená.

- Preklad

Vytvorený, čiže pôvodný dokument a ďalšie dve verzie boli v slovenskom jazyku. Avšak aby bol tento pokus zaujímavejší, som nechal tretiu verziu preložiť do českého jazyka, čím som chcel systém Theses preskúšať v tom ako reaguje na tieto dva, veľmi podobné jazyky. Spomenúť však ešte musím, že v tejto tretej verzii bol zmenený iba jazyk dokumentu. Poradia slov zmenené neboli a tak isto sa nenarábalo ani so synonymami. Akurát v pár vetách bolo potrebné doplniť alebo pozmeniť jedno slovo, ale to len z toho dôvodu, aby boli vety gramaticky správne.

Výsledok tohto malého pokusu bol, teda minimálne pre mňa, celkom prekvapujúci. Čo sa týka doslovného kopírovania, čiže využitia funkcie ctrl+c a ctrl+v, je systém Theses v podstate nepriestrelný. Takýto typ plagiátorstva spoznal, presne tak ako som od neho očakával a pôvodný dokument svietil pekne na červeno. Avšak o ďalších typoch sa to povedať vôbec nedá. Kombinácia zmeny slovosledu a použitia synoným robí systému očividne neprehliadnuteľné problémy a jednoducho si s touto kombináciou nevie rady. V tomto prípade text nesvietil skoro vôbec až na pár výnimiek. Avšak výnimkami bolo vo väčšine len tých pár slov, ktorých pozícia nebola

zmenená a neboli použité ani ich synonymá. Ďalej čo sa týka prekladu do českého jazyka. V podstate môžem kľudne napísať, že systém Theses porovnáva český a slovenský jazyk takým spôsobom, ako kebyže porovnáva dva úplne cudzie jazyky, ktoré spolu nemajú spoločné ani to najmenšie, čo vôbec nie je pravda.

Ako zhrnutie by som napísal, že systém Theses je používaný celkom slušným počtom vysokých škôl. Avšak, podobné dokumenty hľadá a porovnáva len vo svojej databáze. Jedným zo záporov je, že nespolupracuje s inými databázami, čo by ponúklo väčšie množstvo dokumentov, ktoré by bolo možné prehľadávať. Ďalší zápor je nevyhľadávanie podobných dokumentov na internete, aby systém preskúmal aj internetovú stránku, tak je potrebné ju ručne vložiť do systému. Za posledný zápor by som označil skutočnosť, že systém Theses vlastne nepozná podobnosti v slovenskom a českom jazyku a berie ich ako dva úplne odlišné jazyky. Označuje len tie slová, ktoré sa píšu rovnako v oboch jazykoch a čo sa týka slov, ktoré sa líšia len v jednom mäkčeni alebo dĺžni systém neoznačí.

Na záver by som napísal už len toľko, že systém Theses sa snaží bojovať proti plagiátorstvu, ktoré je v dnešnej dobe široko rozšírené a s dnešným rýchlym pokrokom sa v blízkej budúcnosti možno aj dočkáme jeho vylepšenia. Či už čo sa vnímania blízkych jazykov týka alebo kvalitnejšieho detekovania zmien slovosledu a použitia synoným.

Bibliografia

1. Infogram portál pro podporu informační gramotnosti. [Online] [Citace: 19. Marec 2014.] http://www.infogram.cz/findInSection.do?sectionId=1115&categoryId=1168.

2. České vysoké učení technické v Praze. *Ústřední knihovna.* [Online] [Citace: 19. Marec 2014.] http://knihovna.cvut.cz/studium/jak-psat-vskp/doporuceni/plagiatorstvi/co-je-plagiatorstvi.html.

3. Merriam-webster. [Online] [Citace: 19. Marec 2014.] http://www.merriam-webster.com/dictionary/compilation.

4. plagiarism.org. *Glossary.* [Online] [Citace: 19. Marec 2014.] http://www.plagiarism.org/plagiarism-101/glossary/.

5. The JAMA network. [Online] [Citace: 19. Marec 2014.] http://archsurg.jamanetwork.com/article.aspx?articleid=397351.

6. The British Journal of psychiatry. [Online] [Citace: 19. Marec 2014.] http://bjp.rcpsych.org/content/111/480/1111.short.

7. *Zákon č. 618/2003 Z.z., zákon o autorskom práve a právach súvisiacich s autorským právom.*

8. *Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů.*

9. *Zákon č. 131/2002 Z.z., o vysokých školách.*

10. *Zákon č. 111/1998 Sb., o vysokých školách.*

11. **Barnbaum, Dr. C.** *PLAGIARISM: A Student's Guide to Recognizing It and Avoiding It.* [Online] 2002. [Citace: 23. Marec 2014.]

12. Mazda. *Prehľad.* [Online] [Citace: 23. Marec 2014.] http://www.mazda.sk/nase-modely/mx-5-soft-top/.

13. Odtlačok.sk. [Online] [Citace: 23. Marec 2014.] http://www.odtlacok.sk/.

14. bux.sk Knihy, ktorými žijete. *Ukážka z knihy Inferno - Peklo.* [Online] [Citace: 23. 3 2014.] http://www.bux.sk/knihy/nahlad/205708-inferno-peklo.html?page=0.

15. plagiarism.org. *Types of plagiarism.* [Online] [Citace: 24. 3 2014.] http://www.plagiarism.org/plagiarism-101/types-of-plagiarism.

16. University of Alberta. *Why students plagiarize.* [Online] [Citace: 28. 3 2014.] http://guides.library.ualberta.ca/content.php?pid=560222&sid=4618010.

17. **Synek M., Sedláčková H., Vávrová H.** *Jak psát bakalářské, diplomové, doktorské a jiné písemné práce.* Praha : Oeconomica, 2007. ISBN 978-80-245-1212-9.

18. ISO. *About ISO.* [Online] [Citace: 6. Apríl 2014.] http://www.iso.org/iso/home/about.htm.

19. Fakulta baníctva, ekológie, riadenia a geotechnológií. *Predmety.* [Online] [Citace: 11. Apríl 2014.] http://www.fberg.tuke.sk/kg/predmety/SP/STNISO690.doc.

20. http://generator.citace.com/. *Generátor citací.* [Online] [Citace: 12. Apríl 2014.] http://generator.citace.com/.

21. VŠCHT Praha. *Generátor citací vydavatelství VŠCHT Praha.* [Online] [Citace: 13. Apríl 2014.] http://vydavatelstvi.vscht.cz/apps/uid_ea-002/.

22. Portál pro podporu informační gramotnosti. *Odhalování plagiátu.* [Online] [Citace: 14. Apríl 2014.] http://www.infogram.cz/findInSection.do?sectionId=1115&categoryId=1165.

23. turnitin. *About us.* [Online] [Citace: 14. Apríl 2014.] http://www.turnitin.com/en_us/about-us/our-company.

24. iThenticate. *Products.* [Online] [Citace: 14. Apríl 2014.] http://www.ithenticate.com/products.

25. EVE2 Essay Verification Engine:. *What is EVE2?* [Online] [Citace: 15. Apríl 2014.] http://www.canexus.com/abouteve.shtml.

26. The Plagiarism Resource Site. *Software to Detect Plagiarism.* [Online] [Citace: 16. Apríl 2014.] http://plagiarism.bloomfieldmedia.com/z-wordpress/software/.

27. The Plagiarism Resource Site. *WCopyfind Software and Instructions.* [Online] [Citace: 16. 4 2014.] http://plagiarism.bloomfieldmedia.com/z-wordpress/software/wcopyfind/.

28. Viper The Anti-plagiarism Scanner. *Home.* [Online] [Citace: 16. Apríl 2014.] http://www.scanmyessay.com/.

29. Vysokoškolské kvalifikační práce. *Theses.cz.* [Online] [Citace: 16. Apríl 2014.] http://theses.cz/.

30. Informační systém Masarykovej univerzity. *Články.* [Online] [Citace: 26. Apríl 2014.] is.muni.cz/clanky/2009_inforum_cena.pl.

31. **Tim Falla, Paul A. Davies.** *Solutions Pre-Intermediate Student's Book.* Druhé vydanie. místo neznámé : Oxford University Press, 2012. str. 136. 9780194552875.

32. Česko-Slovenská filmová databáze. [Online] [Citace: 11. Apríl 2014.] www.csfd.cz.

Zoznam obrázkov a tabuliek

Obrázok 1: Karta referencie 12
Obrázok 2: Vytvorenie nového zdroja 13
Obrázok 3: Vytvorenie citáte cez citace.com 13
Obrázok 4: Údaje pre novú citáciu 14
Obrázok 5: Úvodná strana generátoru citácií 15
Obrázok 6: Generátor citácií VŠCHT Praha 15
Obrázok 7: Prihlasovanie do systému 21
Obrázok 8: Titulná strana 22
Obrázok 9: Možnosti úschovne 22
Obrázok 10: Úschovna - pridanie nového dokumentu 23
Obrázok 11: Úschovna - Označenie a hľadanie podobností 23
Obrázok 12: Nájdené podobné dokumenty 24
Obrázok 13: Podobnosti textu 25
Obrázok 14: Text z wikipédie 25
Obrázok 15: Podobnosti textu po úprave 26
Obrázok 16: Nájdené podobnosti v dokumente č. 2 27
Obrázok 17: Nájdené podobnosti v dokumente č. 3 28
Obrázok 18: Nájdené podobnosti v dokumente č. 4 29
Obrázok 19: Ďalšie podobné dokumenty 30
Obrázok 20: Podobnosti v záverečnej práci 31

Tabuľka 1: Porovnanie Viper a WriteCheck 19

Prílohy

Príloha č. 1: Dokument Spam

Spam je nevyžiadaná a hromadne rozosielaná správa prakticky rovnakého obsahu. Ide o zneužívanie elektronickej komunikácie, najmä e-mailu. Zväčša je používaný ako reklama, hoci za krátku históriu elektronickej komunikácie bol spam použitý aj z iných dôvodov. Existuje veľa rôznych médií, ktoré sú spamermi zneužívané. Môže to byť napríklad spomínaný e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy, ...

Rozlišujeme dva druhy spamu - prvý je poslanie správy do množstva diskusných skupín (tzv. fórum) Tieto správy sú väčšinou zamerané na distribúciu pornografického materiálu, resp. na propagáciu určitých produktov, napr. z oblasti farmácie. Správa väčšinou nemá žiaden súvis s témou diskusnej skupiny. Je zameraný na ľudí, ktorí často čítajú správy, ale neposkytujú svoje emailové adresy. Druhý typ je emailový spam, zaslaný konkrétnemu človeku priamo na jeho emailovú adresu. Tieto adresy sa získavajú často prehľadávaním diskusných skupín alebo prehľadávaním webových stránok. Emailové adresy získavajú internetové roboty prehľadávaním stránok. Ďalším spôsobom ako získavajú adresy je vyplnenie formulára ponúkajúceho na prvý pohľad zaujímavú protihodnotu. Existujú špecializované firmy alebo jedinci, ktorí vytvárajú databázy adries a tie následne predávajú.

Pretože množstvo diskusných skupín je dnes moderovaných a prevádzkovateľ limituje prevádzku, pokúšajú sa spameri dostať cez automatizované nástroje do čo najväčšieho počtu takýchto diskusných skupín. Na zabránenie masového prístupu sa v súčasnosti používa systém CAPTCHA. Ochrana proti spamu je komplikovaná. Problém je v tom, že neexistuje ostrá hranica medzi spamom a užitočnými emailami a neexistujú príznaky, ktoré by jednoznačne indikovali, že ide o spam. Následkom toho nie je možné vytvoriť prostriedok, ktorý by na sto percent likvidoval spam a na sto percent prepúšťal ostatné emaily.

Príloha č. 2: Druhá verzia dokumentu Spam

Spam je formou nevyžiadanej a hromadne rozosielanej správy s takmer rovnakým jadrom. Jedná sa hlavne o zneužitie formy el. komunikácie, zväčša mailu. Najčastejšie je používaný formou reklamy, ale počas nedlhej histórie el. komunikácie bol používaný aj pre iné účely. Je veľa rozličných druhov médií, ktoré spameri využívajú. Najčastejšie je to mail, rôzne chaty (ako MSN messenger), sms-ky. A skupiny usenet.

Poznáme 2 druhy spamu – tým prvým je posielanie spamu do rôznych fór. Takéto rozosielanie je zväčša zacielené na šírenie pornografického materiálu, určitých produktov, napr. z oblasti lekárnictva. Správa nijak nesúvisí s témou skupiny diskusných fór. Zameriava sa na ľudí ktorí často čítajú správy, ale svoje emaily neposkytujú. Druhým typom je mailový spam, ktorý sa posiela konkrétnej osobe na mailovú adresu. Tieto adresy sa dajú získať prehľadávaním diskusných skupín a webových stránok. Hľadajú ich taktiež internetové roboty ktoré prehľadávajú stránky. Ďalším zo spôsobov je ponúknutie osobe vyplniť formulár za zaujímavú protihodnotu. Taktiež jestvujú špeciálne firmy alebo jednotlivci, ktoré vytvárajú databázy mailových adries a potom ich predávajú.

Keďže v súčasnosti je väčšina diskusných skupín moderovaných, a ich prevádzkovateľ limituje prevádzku, spameri sa snažia dostať cez automatizované nástroje do čo najväčšieho počtu diskusných fór. Systém CAPTCHA bol vyvinutý práve na zabránenie takýchto útokov. Ochrana proti spamu je dosť zložitá. Problém nastáva v tom, že nevieme presne určiť konkrétnu hranicu medzi spamom a užitočnými emailmi, a nie sú žiadne ukazovatele ktoré by jasne stanovil, že sa jedná o spam. V dôsledku toho je nemožné vytvoriť nástroj, ktorý by na 100% likvidoval spam, a ostatné emaily púšťal.

Príloha č. 3: Tretia verzia dokumentu Spam

Hromadne rozosielaná a nevyžiadaná správa s takmer rovnakým obsahom sa nazýva Spam. Jedná sa pritom o zneužitie elektronickej komunikácie, zväčša e-mailu. Spam sa najčastejšie používa ako typ reklamy , aj keď počas krátkodobej histórie elektronickej komunikácie sa Spam použil aj z rozličných dôvodov. Spameri zneužívajú veľa rôznych médií, sú to napríklad vyššie uvedený e-mail, instantné zasielanie správ (napríklad ICQ), skupiny Usenet, krátke textové správy.

Spam sa rozdeľuje na 2 druhy – prvý sa týka posielania správ do mnoho diskusných skupín (tzv. fórum). Hlavnou úlohou takýchto správ je zameriavať sa na distribúciu pornografických materiálov poprípade na propagáciu určitých produktov napr. z oblasti farmácie a medicíny. Správa obvykle nesúvisí s témou danej diskusnej skupiny. Tento druh spamu sa zameriava na skupiny užívateľov, ktorí čítajú svoje správy pravidelne, ale neposkytujú svoje emailové adresy. Ďalším druhom spamu je emailový spam, ktorý sa zasiela konkrétnemu užívateľovi na jeho emailovú adresu. Najčastejšie sú adresy získavané podrobným prehliadaním diskusných fór a webových stránok. Vyplnením formulára, ktorý ponúka na prvý pohľad zaujímavú ponuku, no v skutočnosti je to presný opak je ďalší spôsob ako získať emailovú adresu. Ďalej existujú špecializované spoločnosti a jedinci, ktorí vytvoria databázu adries a tie následne posúvajú a predávajú ďalej. Keďže prevádzkovatelia diskusných fór limitujú prevádzku a fóra sú moderované, spameri sa skúšajú dostať pomocou automatizovaných nástrojov do čo najviac diskusných skupín. V súčasnosti na zabránenie masového prístupu sa používa systém CAPTCHA. Antispamová ochrana je veľmi zložitá. Hlavný problém tkvie v tom, že nie je presne definovaná hranica medzi spamom a užitočnými emailami a taktiež neexistujú príznaky, ktoré by naznačovali, že sa jedná o spam. V dôsledku toho sa nedá vytvoriť ochranno-technologický prostriedok, ktorý by stopercentne likvidoval spam a zároveň prepúšťal ostatné emaily.

Príloha č. 4: Česká verzia dokumentu Spam

Spam je nevyžádaná a hromadně rozesílaná zpráva prakticky stejného obsahu. Jedná se o zneužívání elektronické komunikace, především e-mailu. Spam je nejvíce používaný jako reklama, avšak za krátkou historii elektronické komunikace byl spam použitý i z jiných důvodů. Existuje mnoho různých médií, které jsou spamermi zneužívané. Může to být například již zmiňovaný e-mail, zasílání zpráv (například ICQ), skupiny Usenet, krátké textové zprávy, aj.

Rozlišujeme dva druhy spamu. První je posílání zprávy do většího množství diskuzních skupin (tzv. fórum). Tyto zprávy jsou většinou zaměřené na distribuci pornografického materiálu, respektive na propagaci určitých produktů, např. z oblasti farmacie. Posílaná zpráva většinou nemá žádnou souvislost s tématem diskuzní skupiny. Spam je zaměřený na lidi, kteří často čtou zprávy, ale neposkytují své emailové adresy. Druhý typ je emailový spam, zaslaný konkrétnímu člověku přímo na jeho emailovou adresu. Tyto adresy se získávají hlavně prohledáváním stránek. Emailové adresy získávají internetoví roboti prohledáváním stránek. Dalším způsobem jakým získávají adresy je vyplnění formuláře nabízejícího na první pohled zajímavou protihodnotu.

Existují specializované firmy nebo jedinci, kteří vytvářejí databázi adres a ty následně prodávají. Protože množství diskuzních skupin je dnes moderovaných a provozovatel limituje provoz, pokoušejí se spameři dostat přes automatizované nástroje do co největšího počtu takovýchto diskuzních skupin. Na zabránění masového přístupu se v současnosti používá systém CAPTCHA.

Ochrana proti spamu je komplikovaná. Problém je v tom, že neexistuje ostrá hranice mezi spamem a užitečnými emailami a neexistují příznaky, které by jednoznačně identifikovali, že jde o spam. Následkem toho není možné vytvořit prostředek, který by na sto procent likvidoval spam a na sto procent pouštěl ostatní emaily.