

Čím víc se podvádí, tím víc se systémy na odhalení plagiátů zdokonalují.

Taťána Malá kvůli tomu přišla o místo ministryně spravedlnosti, Petr Krčál o post ministra práce a sociálních věcí. Oba rezignovali, neboť se ocitli v silném podezření, že opisovali ve svých vysokoškolských pracích od jiných autorů a celé dlouhé pasáže vydávali za vlastní. Novináři na to přišli díky antiplagiátorským počítačovým programům, které už v roce 2004 začali vyvíjet odborníci v brněnské Masarykově univerzitě.

Systém se neustále vyvíjí. I s tím, jaké triky a podvody studenti zkoušejí, říká Jitka Brandejsová z Centra výpočetní techniky **Fakulty informatiky** Masarykovy univerzity, která stála u zrodu programů a je členkou týmu, jež je dodnes vylepšuje.

* Dnes existují takové systémy tři, kromě Theses.cz ještě Odevzdej.cz a Repozitar.cz. Jaký je mezi nimi rozdíl?

Theses.cz je výhradně pro vysokoškolské závěrečné práce a absolventské, když jde o vyšší odbornou školu, a do Národního registru vysokoškolských kvalifikačních prací je vkládají školy. Repozitar.cz je určen pro zaměstnance škol a jejich díla, články, publikace, monografie, ale i doktorské habilitační práce. Sem je vkládají školy. Odevzdej.cz slouží k tomu, aby si školy mohly kontrolovat i jiná díla: maturitní práce, referáty, seminární práce. Používají to i střední a vyšší odborné školy a třeba i slovenské ministerstvo práce a sociálních věcí, které si kontroluje projekty.

* Zeptám se rovnou: Opisovala Taťána Malá, nebo ne?

Dívala jsem se v médiích na otisknuté výsledky porovnání její práce ze systému Theses a podobnost kolem pěti procent může být úplně v pohodě, když se všechno řádně odcituje. Pokud se najdou podobnosti zejména u právních diplomek, to neznamená, že autor nutně opisoval, protože se tam často citují zákony. Jenže to, že necitovala, je dost velký prohřešek sám o sobě. V akademické sféře, ve které jsem už patnáct let, nepamatuji doby, že by se podceňovalo citování. Na to se dbalo vždycky.

* Malá plagiátorství odmítá. Ale když se v její diplomce objevila věta, která předtím byla v jiné práci, a v obou jsou dvě stejné gramatické chyby, to nevypadá na náhodu.

To ale musí vždycky prozkoumat odborník. Vždycky. Musí se jasně vyznačit, který dokument vznikl dřív a jestli není ještě třetí původce, odkud se to opisovalo, to všechno může hrát roli. Systém vždy najde jen shodu nebo podobnost, ale to, jestli šlo o plagiát, musí vždy rozhodnout odborník – vedoucí diplomky nebo vedoucí katedry.

* Jak systém funguje? A podle čeho rozhoduje, se kterými texty v obrovském množství akademických prací i článků na internetu bude porovnávat kontrolovaný text?

Pracuje na podobném principu jako vyhledávač Google, ale máme navíc vlastní speciální algoritmus, který nechci prozradit, chráníme si know-how.

* Co systém vlastně porovnává? Věty, odstavce nebo celé pasáže?

Všechno dohromady. Na to jsou speciální algoritmy, které počítají s různými obměnami, jež jsme vynalezli, nějakou dobu nám to trvalo, proto to neprozrazujeme.

* Dříve stačilo přehodit slova ve větě, mírně pozměnit slovosled a systém už podobnost nenašel. Co dnes?

Když přehodíte dvě slova, systém to odhalí. Pokud úplně přeházíte celou větu, stroj to vyhodnotí jako jiný význam a podobnost nenajde. Ale neustále ho zdokonalujeme a i to se v budoucnu vyřeší.

* A stačí nahradit jedno slovo ve větě synonymem?

Když z porovnávané věty vypadne jedno slovo, ale zůstane zbývajících pět, deset, dvacet slov, stroj to najde a označí červeně jako podobnost. A hned to trkne do očí. Existují různé pokusy, jak systém obejít, a proti nim jsme vyvinuli další ochranu. Jeden ze studentů třeba překlopil cizí text do svého a místo mezer mezi slovy psal písmeno A v bílé barvě. Opticky to vypadalo jako mezera, ale ve skutečnosti šlo o řetězec znaků za sebou bez mezer.

* Co na to váš program?

Na to jsme vyvinuli fintu a stroj to dnes vyhodnotí jako jedno dlouhé nesmyslné slovo. Označí celý text, protože není logické, aby existovalo tak dlouhé slovo několikrát za sebou. A následně systém vyše upozornění na pochybný dokument. Jiný student zkoušel nahradit písmeno A podobným znakem v azbuce, aby zmátl program znakem jiné abecedy. I to dnes systém odhalí.

* To opravdu někdo zkoušel?

Ano a kupodivu to byli většinou studenti přicházející z Východu a studující na českých školách. Vždycky se na to přišlo náhodou a už jsme udělali opatření.

* Jak program odliší pasáže, které student převezme z jiné práce, ale řádně je odcituje, od těch, které jen opíše a vydá za vlastní?

To program neodliší a my ani nechceme, aby to odlišoval. Protože navzdory normám různí studenti citují různými způsoby. Kdybychom neoznačili všechny shodné texty, mohlo by se zdát, že ten systém nefunguje.

* A co plagiátorství myšlenek? Tedy, když někdo svými slovy popíše cizí nápad. Může počítačový program odhalit inspiraci?

V odborných oblastech to odhalí už dnes, protože klíčová slova budou hodně podobná. Myslím, že systém odhalí zdroj a pak je opět na člověku, aby rozhodl, zda byla myšlenka ukradená.

* Na druhou stranu se může stát, že dva lidi nezávisle na sobě přijdou na řešení nějakého problému.

Ano a proto, když se patentují závratné objevy, nesmí být nikde zveřejněny nebo popsány, jinak je patentový úřad nevezme.

* A cizojazyčné texty? V roce 2014 se objevila informace, že tehdejší generální tajemník bavorské CSU Andreas Scheuer ve své rigorózní práci z pražské Univerzity Karlovy o deset let dříve opisoval. Univerzita tenkrát oznámila, že systém na odhalení plagiátů opisování neprokázal. Vy jste tehdy řekla, že odhalit plagiát u cizojazyčné práce je nemožné. Jak se to od té doby posunulo?

Posouvá se to, pořád to ještě není implementované, ale je to jedna z výzev, která by v budoucnu měla přijít. Takových výzev je víc – obrázky, grafy, tabulky, technologie na jejich porovnávání ještě není dokonalá, ale pracuje se na zlepšení.

Jeden student třeba překlopil cizí text do svého a místo mezer mezi slovy psal písmeno A v bílé barvě. Vypadalo to jako mezera, ale šlo o řetězec znaků bez mezer. I to systém odhalil.

Foto popis| Vývojářka Jitka Brandejsová.
Foto autor| Foto: Anna Vavříková, MAFRA

O autorovij| Artur Janoušek, reportér MF DNES