

Antiplagiátorský systém (APS) kontroluje originalitu prací uložených v Centrálním registru závěrečných a kvalifikačních prací (CRZP). APS indexuje práce uložené v CRZP a navíc zdroje z volně dostupného internetu s důrazem na zdroje v slovenském jazyce.

Samotný algoritmus není závislý na jazyku. To se projevuje i aktuálně na výsledcích testování prací v anglickém, maďarském, ale i ukrajinském jazyce. Pro vysokou spolehlivost a přesnost detekce podobnosti v konkrétním jazyku je potřebné udělat i množství nastavení a vytvořit specifické, jazykově závislé datové struktury.

Aktuálně hlavním záměrem APS je zkvalitňování procesu vzdělávání. V APS je však skrytý aj potenciál nástroje podporujícího zkvalitňování procesů ve vědě a výzkumu. Tento potenciál je potřebné využít a rozšířit záběr CRZP a APS i na dokumenty neohraňované díkci vysokoškolského zákona (závěrečné a kvalifikační práce) a pojmout tam dokumenty vypracované celou vědeckou komunitou na Slovensku.

Stanovený cíl je možné dosáhnout v těchto hlavních krocích:

- Vypracování studie
 - ❖ o aktuálním stavu zpracování závěrečných a kvalifikačních prací na SAV s cíli
 1. analýza aktuálního stavu úložišť
 2. návrh na unifikaci úložišť
 3. návrh na zapojení úložišť do CRZP
 - ❖ o aktuálním stavu publikační a ediční činnosti SAV a akademických pracovišť v SR s cíli
 1. analýza možností poskytovat služby APS pro účely kontrolu publikační činnosti (s garancí nezveřejňování)
 2. realizovatelnost rozšíření CRZP a APS o práce z publikační a ediční činnosti SAV a akademických pracovišť v SR
 - ❖ o legislativních aspektech předkládaného řešení
 - ❖ o potřebě analytických a programátorských prací
 - ❖ o potřebě finančních zdrojů
 - ❖ o potřebě lidských zdrojů
 - ❖ o časové náročnosti
- Implementace modifikovaného CRZP a APS a jeho pomocných systémů
- Vypracování dokumentace
- Pilotní provoz modifikovaného CRZP a APS a jeho pomocných systémů
- Rutinní provoz modifikovaného CRZP a APS a jeho pomocných systémů

Hlubším ponořením do problematiky detekce plagiátů jsme zjistili, že kvalitu procesu detekce dokáží pozitivně ovlivnit:

- Implementace slovníku synonym slovenského jazyka do tezauru APS
- Implementace kořenových morfémů slovenského jazyka do tezauru APS
- Implementace „stop slov“ slovenského jazyka do tezauru APS
- Implementace „čištění“ korpusu slov slovenského jazyka (vyřazování nesprávných, zkomolených slov)
- Implementace slovenského WordNetu
- Budování korpusu anglického jazyka

Rutinní provoz APS s rozšířenými detekčními schopnostmi nad závěrečnými a kvalifikačními pracemi a nad pracemi publikační a ediční činnosti SAV a akademických pracovišť v SR by mohla být cílovým stavem v období pro nejbližší roky.

V úlohách rozšíření rozsahu testovaných dokumentů je možné postupovat dvěma směry. Prvním směrem je rozšiřování záběru CRZP o dokumenty VaV komunity, které v CRZP zatím nejsou podchycené. Druhým směrem je rozšiřování indexování internetu i o specifické anebo běžně nedostupné elektronické zdroje.

V oblasti rozšíření CRZP je možné uvažovat o těchto okruzích dokumentů:

- Zapojení ústavů SAV do procesu běžného testování závěrečných a kvalifikačních prací, které sú produkované na pracovištích SAV
 - vyžaduje analýzu existujícího stavu s evidencí prací na SAV
 - může vyžadovat vybudování úložiště EZP na SAV
- Umožnit VŠ a SAV dobrovolné testování elektronických dokumentů (jiných jako závěrečné a kvalifikační práce) které vznikají na jejich půdě, s cílem zvýšit jejich ochranu a kvalitu

Když hovoříme o dokumentech, které vznikají jako výsledky vědecké činnosti, máme na mysli především tyto typy:

- dokumenty, které jsou výstupem publikační činnosti (články v sbornících a časopisech, ...)
- dokumenty, které jsou výstupem ediční činnosti (učebnice, skripta, ...)

Specificky zajímavou oblastí použití APS je jeho zapojení do procesu recenzování publikací. V této souvislosti by bylo možné APS poskytnout na použití vydavatelům časopisů a sborníků a výměnou získávat plné texty. Hlavní přínosy spočívají:

- v rozšíření možností recenzenta pro napsání objektivní recenze

- v ochraně autorů, kterých příspěvky jsou po indexování chráněné nejen právně ale i fakticky, protože se dostávají do porovnávání se všemi dokumenty, které přijdu do APS po vytisknutí daného článku
- systémy APS, minimálně jeho prezentační části je také možné použít i pro zveřejňování publikací autorů, kteří by projevíli zájem a publikace nejsou nijak licenčně vázané a tím zabezpečit jejich vyšší dostupnost a citovanost ve vědecké komunitě.

Analýza a implementace slovníků slovenského jazyka je důležitá pro účely zvyšování spolehlivosti detekce podobných textů. Stěžuje mechanické přebírání cizích myšlenek do textu. Budování systémů speciálních slov musí být systematické, aby pokrylo co největší slovní zásobu jazyka. Z korpusu dokumentů CRZP byly vypracované seznamy stop slov pro SJ a AJ. Dokumenty, které vstupují do APS, obsahují různá slova, mnohé překlady anebo sekvence písmen bez významu. Strom slov je z těchto důvodů velmi důležité čistit. Zároveň je možné využít významové závislosti slov. Kromě krátkých a často používaných slov, která nemají reálný vliv na výsledky zjišťování podobností (tzv. stop slova – anebo slovník neindexovaných slov) je možné využít slovníky morfém a synonym. V této souvislosti je možné definovat tyto částečné úlohy:

- Analýza a implementace konverzního algoritmu pro převod slovníku morfém do APS
- Analýza a implementace konverzního algoritmu pro převod slovníku synonym do APS
- Analýza a implementace konverzního algoritmu pro převod slovníku WordNet pro AJ do APS
- Práce ručního dočistění stromu slov a získání relevantního WordNETu pro SJ
- Implementace notificačního mechanismu pro čištění „nových slov“ které jsou indexované už po procesu prvotního očištění slovníku

Případy plagiátorství se nevyhýbají žádné vrstvě společnosti a dnes ještě stále nejsou dostatečně pranýřované a postihované. Ani případy plagiátorství mezi politiky či poslanci nejsou výjimkou. Potřebujeme implementaci etických standardů dostat na podstatně vyšší úroveň. Velmi důležitou oblastí je aj nadále zvyšování povědomí o etice, plagiátorství, o autorských právech, o duševním vlastnictví. Významnou úlohu při zvyšování povědomí musí sehrát rodina, škola, pracoviště a všechny druhy médií. Nejsme úplně na začátku, děti, mládež a dospělí si plně uvědomují, co je krádež (hmotná). Potřebujeme jim však dostat do povědomí i krádež jiného druhu – krádež idejí, textů, obrázků, ... a seznámit je se základními etickými principy. Nebude to lehká úloha, protože v dnešním internetovém světě převládá u dětí a mládeže a části dospělé populace názor, že co je na internetu, to je i moje. Na tomto poli je ještě potřebné vykonat mnoho práce.

K zvýšení povědomí o plagiátorství přispělo už samotné oznámení o tom, že všechny vysoké školy na Slovensku budou povinně využívat systém na odhalování plagiátů. Až po tomto oznámení začal vysoce významně stoupat počet vyhledávání tohoto pojmu na internetu a tím se veřejnost – minimálně studentská – blíže seznamovala s významem tohoto pojmu. Kromě internetu stoupl i počet článků v médiích (aj tištěných) o tomto pojmu, ale ne až tak výrazně.

Existují však skupiny lidí, kteří nehledají poctivé cesty k získání vysokoškolských titulů a i skupiny, které poskytují služby těmto skupinám. Není složité za několik minut najít několik internetových stránek poskytujících vypracování podkladů k různým druhům prací.

Každá vysoká škola má připravenou vlastní směrnici týkající se závěrečných a kvalifikačních prací, kterou si připravila na základě všeobecné směrnice, kterou vydalo MŠVVaŠ SR. Dodržování směrnic týkajících se postupů při přípravě, odevzdávání a vyhodnocování protokolů o kontrole originality přispívá k zvýšení kvality odevzdávaných prací.

Od 1. 9. 2011 jsou práce přístupné na portále CRZP (ty, které došly do CRZP po 31. 8. 2011) a jsou tedy nejen pod kontrolou vysokých škol, ale i pod kontrolou veřejnosti.