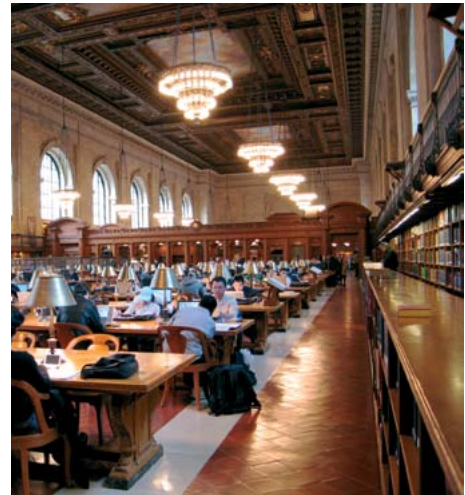


## Systém na odhaľovanie plagiátov

Algoritmus pre kontrolu plagiátov je vyvíjaný od roku 2008. V roku 2010 bol ANTIPLAG, ako sa subsystém pracovne volá, implementovaný ako súčasť projektu Centra vedecko-technických informácií SR (CVTI SR) pod záštitou Ministerstva školstva SR. Oficiálny názov tohto projektu je **Systém na odhaľovanie plagiátov pre potreby slovenských akademických a výskumných inštitúcií** a jeho fyzická realizácia vo forme portálu má názov **Centrálny register záverečných a kvalifikačných prác** (dostupný na [www.crzp.sk](http://www.crzp.sk)). V budúcnosti je možné uvažovať aj nad využitím systému pri vedeckých a odborných publikáciách a to pri porovnaní vlastného autorského prínosu voči iným autorom ako aj voči autorovým predchádzajúcim prácam. Systém je využiteľný aj pre grantové výskumné projekty.



Systém ANTIPLAG je systém pre podporu rozhodovania. Jeho výstupom je protokol, elektronický dokument, ktorý upozorňuje na dokumenty, v ktorých bolo identifikované podozrenie na plagiátorstvo. Je nástrojom, ktorý upozorňuje na zhodné alebo identické dokumenty z internetu alebo práce z iných pracovísk, ktoré mohli uniknúť pozornosti školiteľa alebo oponenta pri posudzovaní záverečných prác.

**Plagiátorstvo** je definované ako kopírovanie alebo používanie cudzích myšlienok a ich prezentácia spôsobom, ktorý vytvára dojem, že ide o originálne myšlienky plagiátora. Teda ide najmä o používanie výsledkov práce iného autora bez jeho citovania.



Cieľom algoritmu je pomáhať **identifikovať podozrenia** a potláčať rôzne formy plagiátorstva. Vďaka plošnému nasadeniu a premyslenému indexovaniu „živého“ internetu je možné zachytiť mnohé pokusy kopírovať práce z iných škôl, predchádzajúcich rokov alebo z internetu v domnení, že množstvo zdrojov je také veľké, že nie je možné prehľadať a porovnať všetky. Systém v sebe **integruje niekoľko prvkov**, akými sú vyhľadávanie vo vlastnej databáze, distribuovanú architektúru dátových úložísk, indexáciu dát, aktívne vyhľadávanie (web crawler), algoritmy pre porovnávanie textu, a iné. Použitá technológia **využíva viacstupňový mechanizmus kontroly práce voči korpusu (indexu) prác**.

V základných bodoch je možné algoritmus charakterizovať takto:

1. Vytvorenie **digitálneho odtlačku** dokumentu pomocou špecializovaných nástrojov so zapojením stromu pojmov
2. **Porovnanie odtlačku** voči dokumentom v korpuse
3. Korpus je pravidelne **doplňaný dátami**, ktoré dodávajú školy, ale aj pomocou automatizovaných webových prehľadávačov z internetu
4. **Vo finálnej fáze sa produkuje protokol** s vyznačením bibliografických citácií a liniek na podozrivé dokumenty s informáciami o zhode a pravdepodobnosti zachytenej podobnosti