Juraj Holas

Návrh projektu

## Téma:

Odhadovanie počtu a štruktúry zamestnancov v oblasti výskumu a vývoja na základe celkového počtu obyvateľov a výdavkov v tejto oblasti.

## Popis:

Projekt sa zameria na odhadovanie ročných štatistík zamestnancov v oblasti výskumu a vývoja v SR. Skúmaný bude nielen celkový počet týchto zamestnancov, ale aj ich demografická štruktúra, menovite zastúpenie pohlaví, najvyššie dosiahnuté vzdelanie a pracovná pozícia. Zisťovať sa bude vzťah týchto parametrov ku celkovému počtu a štruktúre obyvateľstva SR, a tiež ku objemu a charakteru financií využitých v tomto sektore.

## Zdroje dát:

Všetky potrebné dáta budú získavané zo stránky Štatistického úradu Slovenskej republiky - <http://slovak.statistics.sk/>. Využívané budú údaje z databázy *STATdat*, predovšetkým z nasledujúcich tabuliek:

* Základné charakteristiky obyvateľstva (absolútne údaje) [om0001rs]
* Výdavky na výskum a vývoj [vt0002rs]
* Zamestnanci výskumu a vývoja [vt0001rs]
* Štruktúra zamestnancov výskumu a vývoja (v %) [vt1009rs]

V prípade potreby pribudnú ďalšie tabuľky z uvedenej databázy v priebehu práce na projekte.

## Rozdelenie dát:

Všetky informácie získané z uvedených zdrojov sú rozčlenené podľa rokov, čo sa odzrkadlí aj na ich rozdelení v priebehu projektu. Dáta z posledných troch rokov štatistiky, teda roky 2011 až 2013, budú slúžiť ako testovacie dáta na zisťovanie úspešnosti získaného algoritmu. Všetky skoršie štatistiky, teda roky 2000-2010, budú použité na trénovanie.

## Vyhodnocovanie úspešnosti:

Po získaní požadovaného algoritmu z trénovacích dát sa spočíta jeho úspešnosť vzhľadom ku testovacím dátam. Úspešnosť na každom odhadovanom atribúte bude reprezentovaná ako štvorec chyby na normalizovanej hodnote tohto atribútu. Celková úspešnosť bude následne vyhodnotená ako súčet chýb na jednotlivých atribútoch.