

Hodnocení vedoucího diplomové práce

Autor/autorka práce: **Mario Kamburov**

Název práce: **Optimální metody dataminingu pro zpracování semistrukturovaných medicínských dat**

Cílem diplomové práce byl návrh a implementace zpracování semistrukturovaných medicínských dat. Výsledné řešení mělo umožňovat korekce lékařských textů na základě velmi rozsáhlých a různorodých nestrukturovaných a semistrukturovaných údajů z lékařských zpráv.

V práci jsou teoreticky popsány možnosti zpracování přirozeného jazyka a datamining algoritmy, které by bylo možno použít pro klasifikaci textu. Na základě podrobné analýzy spočívající v porovnávání jednotlivých databází a datamining algoritmů jsou vybrány optimální metody pro rychlou a maximálně bezchybnou korekci medicínských textů.

Před samotnou implementací bylo potřeba připravit dostatečně velké množství dat v podobě slovníků, které by mohly posloužit jako stabilní podklad ke korekci lékařských textů. Data byla připravována a očišťována poloautomaticky v prostředí MS Excell. Výstupem byly CSV soubory vhodné pro import do slovníků, které byly následně využity ke korekci textů a odstranění velmi často používaných zkratk.

Pro účely navrhování oprav zkratk a celkovou korekci lékařských textů je používáno fulltextové vyhledávání s využitím CouchDB a Apache Lucene.

Diplomová práce byla oproti verzi předložené v minulém roce značně přepracována. Diplomant akceptoval všechny připomínky, navržené programové řešení je funkční a použitelné v experimentálním medicínském systému, který je vyvíjen na KIV FAV. Na základě vlastní iniciativy diplomant rozšířil možnosti zpracování. Výsledky byly konzultovány v odborném lékařském prostředí.

Práce doznala značné zlepšení i po formální stránce, přestože po jazykové, především stylistické stránce lze ještě mít určité výhrady.

Zadání bylo splněno.

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 25.8.2016

Doc. Dr. Ing. Jana Klečková

SOUHLASÍ
S OPRACOVÁNÍM

Česká univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky