

# Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Jan Knížek  
**Název práce** Comparison of approaches to text classification  
**Rok odevzdání** 2019  
**Studijní program** Informatika      **Studijní obor** Obecná informatika

**Autor posudku** Jiří Hana      **Role** Vedoucí  
**Pracoviště** ÚFAL

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

## K celé práci

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rozsah práce ... <i>textová i implementační část, zohlednění náročnosti</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komentář	Cílem práce je automaticky poznat, zda slovné hodnocení restaurace je užitečné nebo neužitečné. Na tomto úkolu je porovnáno několik algoritmů a přístupů strojového učení a různé sady features.  Tyto systémy jsou trénovány a testovány na datech poskytnutými serverem Yelp. Hodnocení je považováno za užitečné, pokud ho za takové prohlásili aspoň dva uživatelé serveru. Bohužel žádný z přístupů nedosáhl lepších výsledků než baseline. Jedním z důvodů může být šum v testovacích datech - jak upozorňuje autor práce, mezi skutečnou užitečností hodnocení a označením hodnocení jako užitečným uživateli je jen omezená korelace. Bylo by zajímavé zhodnotit systém na datech ručně anotovaných nestraným anotátorem. To je ovšem samozřejmě mimo rozsah bakalářské práce.			

## Textová část práce

	lepší	OK	horší	nevyhovuje
Formální úprava ... <i>jazyková úroveň, typografická úroveň, citace</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Struktura textu ... <i>kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analýza	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vývojová dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uživatelská dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komentář	Práce je psaná anglicky. Je přehledně a logicky strukturovaná, text je srozumitelný.			

## Implementační část práce

	lepší	OK	horší	nevyhovuje
Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stabilita implementace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komentář	Systém je pěkně navržen a dokumentován. Kromě přílohy k práci je umístěn i na githubu. Lze snadno nainstalovat a pustit.			

**Celkové hodnocení** Výborně  
**Práci navrhuji na zvláštní ocenění** Ne

**Datum** 26.8.2019

**Podpis**