

# HODNOCENÍ BAKALÁŘSKÉ PRÁCE

Vedoucí BP

Jméno bakaláře: Tomáš Smolík

Garantující katedra: KKY

Název bakalářské práce: Shlukování textů podle jejich podobnosti pomocí modulu Scikit-learn

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	Samostatnost zpracování tématu BP	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Vhodnost použitých metod	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Správnost získaných výsledků	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Vlastní přínos	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Doplnění hodnocení, připomínky, dotazy:

Cílem této práce bylo porovnat vhodnost testovaných metod v úloze shlukování novinových článků z internetových zpravodajských serverů. Testována byla shlukovací metoda K-means a metody pro předzpracování dat PCA (analýza hlavních komponent) a LSA (latentní sémantická analýza).

Student v úvodu práce velice pěkně shrnuje teoretické poznatky z oboru klasifikace dat a zejména oblasti shlukování, popisuje jednotlivé použité metody a způsoby vyhodnocení získaných výsledků. Dále se student v práci zabývá popisem tří jím navržených experimentů určených k otestování různých aspektů testovaných metod. V rámci těchto experimentů student důsledně otestoval různá nastavení a kombinace jednotlivých metod.

Výsledky jednotlivých experimentů jsou detailně vyhodnoceny a rozebrány, v závěru práce jsou shrnuty výsledky všech experimentů a na základě těchto výsledků jsou doporučeny metody s nejlepšími výsledky jako nejvhodnější pro daný problém.

### Doplňující dotazy:

- Podle dosažených výsledků vašich experimentů předzpracování dat metodami PCA a LSA nezpůsobilo zlepšení výsledků shlukování. Myslíte, že by metody lépe fungovaly pokud byste zpracovával řádově větší množství dat?
- Píšete, že horší výsledky při použití metody LSA jsou způsobeny také jejím nevhodným použitím. Napadá vás nějaký způsob jak tuto metodu vhodně využít při shlukování?

Splnění bodů zadání	<input checked="" type="checkbox"/> úplně	<input type="checkbox"/> částečně	<input type="checkbox"/> nesplněno	
Doporučení práce k obhajobě	<input checked="" type="checkbox"/> ano	<input type="checkbox"/> ne		
Celkové hodnocení práce	<input checked="" type="checkbox"/> výborně	<input type="checkbox"/> velmi dobře	<input type="checkbox"/> dobře	<input type="checkbox"/> nevyhověl
Jméno, příjmení, titul vedoucího BP: Lucie Skorkovská, Ing.				
Pracoviště vedoucího BP: KKY				

26.8.2013

Datum

Podpis