

POSUDEK PRÁCE

Jméno studenta: Zdeněk Záleský

Název práce: Návrh robota pro automatické vyhledávání informací na Internetu týkajících se zvolených témat a jejich ukládání do databáze

Autor posudku: doc. Dr. Ing. Tomáš Brandejský

Cíl práce: Cílem práce je navrhnout automatizovaný vyhledávač informací v prostředí Internetu vztahujících se ke zvoleným tématům (např. počasí, kulturní a společenské akce, doprava apod.) a jejich ukládání do vhodné databáze pro další zpracování.

Vyhledávač může využívat komunikaci se standardními vyhledávači (Google, Seznam apod.), ale může mít i podobu samostatného web crawleru dle zhodnocení situace studentem.

Cílem je získávat maximálně relevantní informace a ukládat je do vhodné databáze (např. Apache Cassandra) pro další zpracování v prostředí Apache Spark Streaming, nebo Apache Kafka.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)			
	1	2	3	4
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Dílčí připomínky a náměty:

Některé nešikovné formulace, jako

str. 24 „Uvedený obrázek má za následek následující cestu Xpath:“

Chybí detailní popis instalace prostředí (na ne zcela přesně referovaném portálu hub.docker.com je potřebné získat uživatelský účet).

Celkové posouzení práce a zdůvodnění výsledné známky:

I přes výše uvedené drobné nedostatky se jedná o zdařilou práci, kde student prokázal schopnost osvojit si znalosti nad rámec bakalářského studia a orientovat se v novém prostředí.

Vyhodnocení kontroly textu práce pomocí systému pro odhalování plagiátu:

Nejvyšší míra shody 21% byla při kontrole zjištěno u souboru middlewares.py (jednom ze souborů přílohy, který definuje web spidera). Ostatní soubory mají shodu pod 5%. Web spider je založen na unifikovaném prostředí Scrapy, proto je takováto míra shody očekávatelná. Vlastní text práce je rovněž původní.

Otázky k obhajobě:

Vyjádřete se prosím, proč bylo odkazováno na hub.docker.com a kontejnery nebyly součástí práce, stejně jako k výsledku kontroly plagiátorství u souboru middlewares.py.

Práci k obhajobě.

Navržená výsledná známka:

V Pardubicích, dne 5. ledna 2021

podpis

Doc. Dr. Ing. Tomáš Brandejský v.r.