



ZADÁNÍ BAKALÁ SKÉ PRÁCE

Název: Nástroj pro m ení datové kvality pomocí SQL dotaz
Student: Vojt ch Malý
Vedoucí: Ing. Tereza Mlyná ová, Ph.D.
Studijní program: Informatika
Studijní obor: Teoretická informatika
Katedra: Katedra teoretické informatiky
Platnost zadání: Do konce letního semestru 2017/18

Pokyny pro vypracování

- 1) Seznamte se s problémem datové kvality v datovém skladu.
- 2) Seznamte se s obecnou architekturou úloh v Jenkins CI.
- 3) Zhodno te použití technik reaktivního programování nebo knihovny Spring Integration oproti implementaci úloh v Jenkins CI.
- 4) V jazyce Java implementujte nástroj pro spoušt ní úloh m ících datovou kvalitu pomocí SQL dotaz a jejich následné vyhodnocování, p í emž SQL dotazy mohou využívat výsledky jiných SQL dotaz , a tím mohou tvo it orientovaný graf závislostí.
Jednotlivé úlohy eší spoušt ní r zných typ m ení, které slouží pro vyhodnocení stavu datové kvality ve zkoumaném úložišti dat.
- 5) Vyberte grafové algoritmy pro transformaci a procházení orientovanými grafy tak, aby jejich použití bylo použitelné ve vyvinutém nástroji na m ení kvality dat.
- 6) Navrhn te a implementujte algoritmus pro paralelní zpracování jednotlivých úloh vycházející z analýzy z bodu 5).

Seznam odborné literatury

- [1] C.Batini, M. Scannapieca. Data Quality: Concepts, Methodologies and Techniques. Springer, 2006.
- [2] Jenkins. GitHub repository (2016), URL: <https://github.com/jenkinsci/jenkins>
- [3] Jenkins. Documentation (2016), URL: <https://jenkins.io/doc/>
- [4] M. Fisher, J. Partner, M. Bogoevici, I. Fuld. Spring Integration in Action. Manning, 2012.
- [5] Nickolay Tsvetinov. Learning Reactive Programming with Java 8. Packt Publishing, 2015.
- [6] Josef Kolá . Teoretická informatika. eská informatická spole nost, 2004.

doc. Ing. Jan Janoušek, Ph.D.
vedoucí katedry

prof. Ing. Pavel Tvrdík, CSc.
d kan

V Praze dne 3. ledna 2017