

Klasifikace problémů kvality dat

1 Dimenze kvality dat

Systém

- Přístupnost (accessibility)
Jsou data snadno přístupná všem relevantním osobám a lze je v případě potřeby transformovat?
- Zabezpečení přístupu k datům (access security)
Lze přístup k datům omezit pouze na relevantní osoby a systémy?

Využití

- Aktuálnost (timeliness)
Jsou data dostatečně aktuální pro všechny účely?
- Tvorba hodnoty (value-added)
Lze z dat generovat relevantní přímou nebo nepřímou hodnotu pro naši společnost?
- Úplnost (completeness)
Jsou data úplná (z hlediska datových polí)?
- Přiměřený rozsah (appropriate amount of data)
Je k dispozici dostatečné množství dat (z hlediska datových záznamů)?
- Relevance (relevancy)
Jsou k dispozici všechna relevantní data (z hlediska datových polí i datových záznamů)?

Obsah

- Důvěryhodnost (believability)
Jsou data autentická a důvěryhodná?
- Reputace (reputation)
Je zdroj dat důvěryhodný a známý vysokou kvalitou dat?
- Objektivita (objectivity)
Jsou data bez osobních nebo AI-založených hodnocení a zkreslení?
- Správnost (accuracy)
Jsou data správná?

Prezentace

- Přehlednost (concise representation)
Jsou data prezentována stručně a bez dodatečných nerelevantních informací?
- Jednotná prezentace (representational consistency)
Jsou data vždy zobrazována ve stejném formátu?
- Srozumitelnost (ease of understanding)
Jsou data jednoznačná a snadno pochopitelná?
- Interpretovatelnost (interpretability)
Jsou data uvedena ve vhodném jazyce a vhodných jednotkách?

2 Typy dat

Struktura dat

- Strukturovaná data
Data jsou uložena v předem definované struktuře (předdefinovaná datová pole)
- Nestrukturovaná data
Data existují v různých formátech a bez pevné struktury (např. texty, obrázky, videa)
- Polostrukturovaná data
Data jsou částečně strukturovaná a částečně nestrukturovaná

Způsob sběru dat

- Sensorová data
Data byla (automaticky) získána pomocí senzorů (často senzory ve strojích)
- Manuální data
Data byla získána manuálně pozorováním, dotazováním nebo odvozováním
- Systémová data
Data byla vypočítána, agregována, predikována nebo simulována systémem

3 Počet dotčených dat

- jednotlivá data nebo malé množství dat
- data určitého způsobu sběru (např. z konkrétního senzoru)
- data určitého způsobu zpracování (např. data agregovaná systémem)
- data s určitými obsahovými charakteristikami (např. všechna data z prosince)
- všechna data

4 Dopady

- dopad na efektivitu interních procesů a/nebo rozhodování
- dopad na kvalitu interních procesů a/nebo rozhodování
- dopad na efektivitu externích procesů a/nebo rozhodování
- dopad na kvalitu externích procesů a/nebo rozhodování

5 Řešení problémů

- úroveň řešení problému (technická, organizační)
- složitost řešení (jednoduché, komplexní)
- časová rovina (náprava aktuálně existujících problémů kvality, prevence budoucích problémů kvality)