Polosemestrální písemka!

15. listopadu

Bližší informace 1. listopadu

Rozsáhlé databáze osobních informací

Vašek Matyáš

PV080

Agregace dat

- Seskupování (osobních) dat do rozsáhlých databází. Agregace (z angl. *aggregation*).
- Tímto kombinováním dat o určité citlivosti lze získat informace daleko citlivější, které jinak spadají do kategorie s vyššími požadavky na ochranu.

Zákon o ochraně osobních údajů (101/2000 Sb.) – Povinnosti správce

Mj. zákon říká:

- nesdružovat osobní údaje, které byly získány k rozdílným účelům, pokud zvláštní zákon nestanoví jinak

Žadatel o investici

- Chodil roky ke stejnému obvodnímu lékaři.
- Uzavřel před měsícem vysokou živ. pojistku.
- V minulém čtvrtletí byl u specialisty.
- Přede dvěma měsíci změnil obvodního lékaře.

Odvození (Inference, i angl.)

- Odvození informací o vyšší citlivosti zpracováním a analýzou skupiny informací o nižší citlivosti.

nebo

- Nepřímý přístup k informacím bez přímého přístupu k datům, která tyto informace reprezentují.

Příklad politiky klinických IS, British Medical Association

- Musí být zavedena účinná opatření proti agregaci osobních zdravotních informací.
- Pacienti, k jejichž seznamu řízení přístupu má být přidána další osoba, musí být zvlášť upozorněni, pokud již tato osoba má přístup ke zdravotním informacím velkého množství lidí.

Co když máte informace o finanční situaci a zdrav. stavu

1. Přítele/kyně, resp. manžela/ky.
2. Spolupracovníka, nadřízeného...
3. Všech studentů/zaměstnanců FI.
4. Všech obyvatel místa, kde žijete.
5. Všech klientů určité firmy (banky, zdravotní pojišťovny...).
6. Všech (většiny) občanů.

Pravděpodobnost neoprávněného použití

- Počet osob, které mají k informacím přístup (operátoři, uživatelé systému ap.).
- Hodnota informací.
  - Výše trestu těm, kdo data jiných řádně neohlídali a spolupodíleli se tak na jejich úniku.
  - Výše trestu těm, kdo s nimi neoprávněně manipulují.
  - Úroveň ochranných mechanismů.

Řešení?

- U menších souborů osobních dat provádět agregaci jen v nutných případech.
- U větších souborů neprovádět agregaci.

- Statistické databáze!

Statistické databáze

- Obsahují citlivé údaje o jednotlivcích.
- Jejich využití má být **jen** pro statistické dotazy k vytvoření obrazu o celkových potřebách obyvatelstva a formulování (vládní) politiky.
  - podpora církví, regionů/měst atd.
- Výsledky dotazů v takovýchto databázích nesmějí poskytnout údaje o jednotlivcích.

Studium statistických databází

- USA, 70. léta, databáze ze sčítání lidu.
- Dorothy Denning
  - Studium používaných způsobů pro formulaci dotazů a získávání odpovědí.
  - Ty povolovaly (netriviální!) dotazy, které umožnily získat údajně tajné informace o jednotlivci.
  - ☺ Údajně nedůvěra ve zjistění Denningové – dokud nezjistila plat svého šéfa sérií legitimních dotazů.

Příklad kritického dotazu

Kolik je měst s 15-16 000 obyvatel

& s muži, evangelíky, slovenské nár., 36-40 let

& jejich ženy, 28-30 let žijí mimo toto město

& 2 děti do 10 let žijí s těmito ženami

& 1 dítě nad 18 žije s těmito muži

& muž žije ve vlastním domě, plocha nad 200m^2
a domácnost má/používá aspoň 2 automobily.

Protiopatření ve statistických databázích I.

Náhodný výběr

- Každý dotaz je zodpovězen na základě vyhodnocení náhodně vybraných záznamů ze všech existujících záznamů.
- Technika nyní používaná v americké databázi údajů ze sčítání lidu.

Protiopatření ve statistických databázích II.

Minimální rozsah dotazu

- Minimum celkového počtu záznamů použitých pro tvorbu odpovědí.

nebo

- Minimum počtu záznamů použitých pro tvorbu odpovědí na každou část dotazu.

Protiopatření ve statistických databázích III.

Perturbační (zmatečné) techniky

- Přidání pseudonáhodného „šumu“:
  - Odpovědi konzistentní, ale získání spolehlivé odpovědi na sérii podobných dotazů není možné.

1. K záznamům zahrnutým pro vyhodnocení dotazů se přidají další náhodně vybrané podobné záznamy
2. Vypočtená hodnota nebo mezihodnoty jsou zaokrouhlovány nebo mírně pozměněny.

- Podle některých definic zahrnují *náhodný výběr*.

Polosemestrální písemka!

15. listopadu

Bližší informace 1. listopadu