1. Zapytanie bezpieczne vs. niezależne od dziedziny:

$$\{x \mid P(x) \wedge \exists y < 0\}$$

Zapytanie koniunkcyjne

$$Q(x_1, x_2, \dots, x_n) :- P(x_1, x_2), P(x_2, y), \dots, R(y', x_n)$$

P	
	x_1, x_2
	x_2, y
	$\vdots$

T	
	$\vdots$

K	
	y', x_n

baza danych to zb. faktów

$$\{P(x_1, x_2), P(x_2, y), \dots, R(y', x_n)\}$$

Def. Dla danego CQ Q kanoniczną bazą danych dla Q nazywamy bazę D^Q składającą się z faktów odpowiadających atomom Q.

Przykład. $(\exists x \exists y)$ węzły

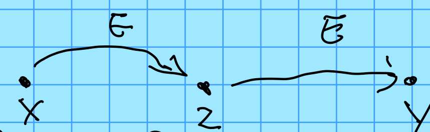
$$Q() :- E(x, 2), E(2, y)$$

1) $D^Q = \{E(x,2), E(2,y)\}$

2)

E	
	$x, 2$
	$2, y$

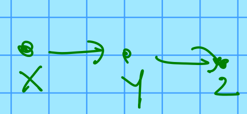
3)



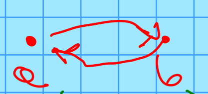
Niech $Q()$ będzie CQ.

Czy $D^Q \models Q$? Oczywiście

Np. kanoniczne zapytanie dla danej bazy — koniunkcja atomów odp. faktom z D



$\not\models$



$$\exists x, y, z \; E(x, y) \wedge E(y, z)$$

Lemat: Niech $Q()$ będzie CQ bez zm. wolnych, a D – baza danych.

Następujące warunki są równoważne:

- $D \models Q$
- Istnieje homomorfizm $h: D^Q \to D$

nie ma kwantyfikatorów

Uwaga $D^Q \neq D$

baza kanoniczna dla Q

jakaś baza

$\{ \varphi(a,d), \varphi(c,a), R(a,a,b), \vdots \}$

Szkic dowodu:

Niech $Q = \exists x_1, \dots, \exists x_n \, \varphi(x_1, \dots, x_n)$

$\Downarrow$) Załóżmy, że $D \models Q$, tzn. że istnieją takie $a_1, \dots, a_n \in \mathrm{adom}(D)$, że $D \models \varphi(\underbrace{a_1, \dots, a_n})$

Wtedy funkcja $h: \{x_1, \dots, x_n\} \to \mathrm{adom}(D)$ zdefiniowana tak:

$\underbrace{h(x_i) = a_i}$ (dla $i = 1, \dots, n$)

jest homomorfizmem.

$\Uparrow$) analogiczny pomysł homomorfizm $\to$ wartościowanie

Ewaluacja zapytań

$\underbrace{D \models Q}$?

(model checking)

Zawieranie się zapytań

$\underbrace{Q \subseteq Q'}$ $\equiv \underbrace{\forall D \;\; D \models Q \Rightarrow D \models Q'}$

Tw. ('77) Niech Q, Q' będą CQ bez zm. wolnych.

Następujące warunki są równoważne

1) $Q \subseteq Q'$
2) Istnieje homomorfizm $h: D^{Q'} \to D^Q$
3) $D^Q \models Q'$

1)⇒2) Zał. $Q \subseteq Q'$. Wiemy, że $D^Q \models Q$. W takim razie $\underbrace{D^Q \models Q'}$

W takim razie jest homomorfizm $h: D^{Q'} \to D^Q$

2)⇒3) Z lematu.

3)⇒1) $D^Q \models Q'$. Z lematu istnieje homomorfizm $h: D^{Q'} \to D^Q$.

Weźmy dowolną bazę danych D.

Załóżmy, że $D \models Q$. Istnieje homomorfizm $h': D^Q \to D$.

Dlaczego $\underbrace{D \models Q'}$???

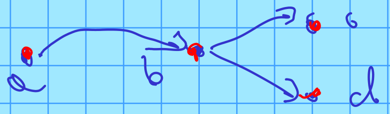
$\Updownarrow$

istnieje $h'': D^{Q'} \to D$

Tak! Zdefiniujmy, że $h''(x) = h'(h(x))$

$h'': D^{Q'} \to D$

jest homomorfizmem. Z lematu $D \models Q'$.

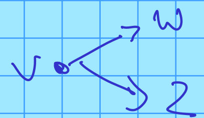
→ Q_1 

$Q_1 :- E(a,b), E(b,c), E(b,d)$

$Q_1 \subseteq Q_2$

$\exists h: D^{Q_2} \to D^{Q_1}$

$E(v,w) \to E(\overset{h(v)}{b}, \overset{h(w)}{c})$

→ Q_2: 

$E(v,w), E(v,z)$

$h(v) = b$
$h(w) = c$
$h(z) = d$

$\{ \text{Osoba}(i,n) \wedge \text{Wpis}(i,p,o) \}$

$\{ \text{Wpis}(i,p,o) \}$

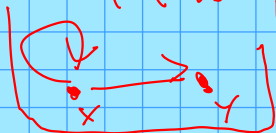
$Q_3 \subseteq Q_1$

$h: D^{Q_1} \to D^{Q_3}$

$Q_3 :- E(x,x), E(x,y)$

$h(a) = h(b) = h(c) = x$

$h(d) = y$

→ Q_3 

Q_4: $p \to r \to s$

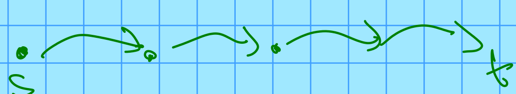
? $Q_1 \subseteq Q_4$? ✓

? $Q_4 \subseteq Q_1$?

$Q_1 \equiv Q_4$

Jakie problemy z wyrażalnością zapytań koniunkcyjnych? algebry relacji?

$E(S,T)$



$Q_4(s,t) :- E(s,t_1), E(t_1,t_2), E(t_2,t_3), E(t_3,t)$

1) $Q_*(s,t)$ — "istnieje ścieżka dow. długości z s do t"

2) mamy relację "drzewo" chcielibyśmy zdef. rel. "przodek"

3) przechodnie domknięcie

Te rzeczy nie są wyrażalne w prd/rrk, ani w algebrze relacji

gry Ehrenfeuchta-Fraïssego — narzędzie do pokazywania takich wyników

Datalog — język, w którym umożliwiamy użycie rekurencji.

Program w datalogu to zb. reguł postaci

$T(x_1, \ldots, x_k) :- R_1(\bar{u}_1), \ldots, R_m(\bar{u}_n)$ ← tak jak CQ, ale T może wystąpić jako jeden z R_i

Symbole relacyjne → intensjonalne (rekurencyjne) / ekstensjonalne (te z bazy danych)

1) Przykład (domknięcie przechodnie):

$$T(x,y) :- E(x,y)$$
$$T(x,y) :- E(x,z),\ T(z,y)$$

$$T = E \cup \pi_{1,4}(\sigma_{2=3}(E \times T))$$

2)

$$odd(x,y) :- E(x,y)$$
$$odd(x,y) :- E(x,z),\ even(z,y)$$
$$even(x,y) :- E(x,z),\ odd(z,y)$$

$$\begin{cases} odd = E \cup \pi_{1,4}(\sigma_{2=3}(E \times even)) \\ even = \pi_{1,4}(\sigma_{2=3}(E \times odd)) \end{cases}$$

Semantyka programu datalogowego:

- najmniejsze rozw. równania
- zdef. sposób obliczania wyniku

Szukamy rozwiązania takiego układu równań rel.

Propozycja: weź $T = adom(D) \times adom(D)$?

"dobre" rozw. ale bezużyteczne

To o co nam chodzi to nie dowolne rozwiązanie ale najmniejsze.

Tw. Każdy taki układ ma najmniejsze rozwiązanie (Knastera-Tarskiego)

Algorytm ewaluacji programu π

1. ustaw wszystkie intensjonalne relacje na $\emptyset$
2. zastosuj wszystkie reguły z π równolegle (tzn. wylicz nowe wersje wszystkich tych relacji)
3. powtarzaj 2. tak długo aż osiągniesz punkt stały (relacje przestaną się zmieniać)
4. Zwróć wynik

$$T(x,y) :- E(x,y)$$
$$T(x,y) :- E(x,z),\ T(z,y)$$

$$T^0 = \emptyset$$
$$T^1 = E \cup \pi_{1,4}(\sigma_{2=3}(E \times T^0)) = E$$
$$T^2 = E \cup \pi_{1,4}(\sigma_{2=3}(E \times T^1)) = E \cup \pi_{1,4}(\sigma_{2=3}(E^2))$$

T^3 = ścieżki dług. 1 lub 2 lub 3

⋮

$$T^{n+1} = \{(a,b) : E(a,b) \vee \exists z\,(E(a,z) \wedge T^{n}(z,b))\}$$

Można pokazać, że $T^n = \{(a,b) \mid$ istnieje ścieżka długości $\leq n$ z a do $b\}$

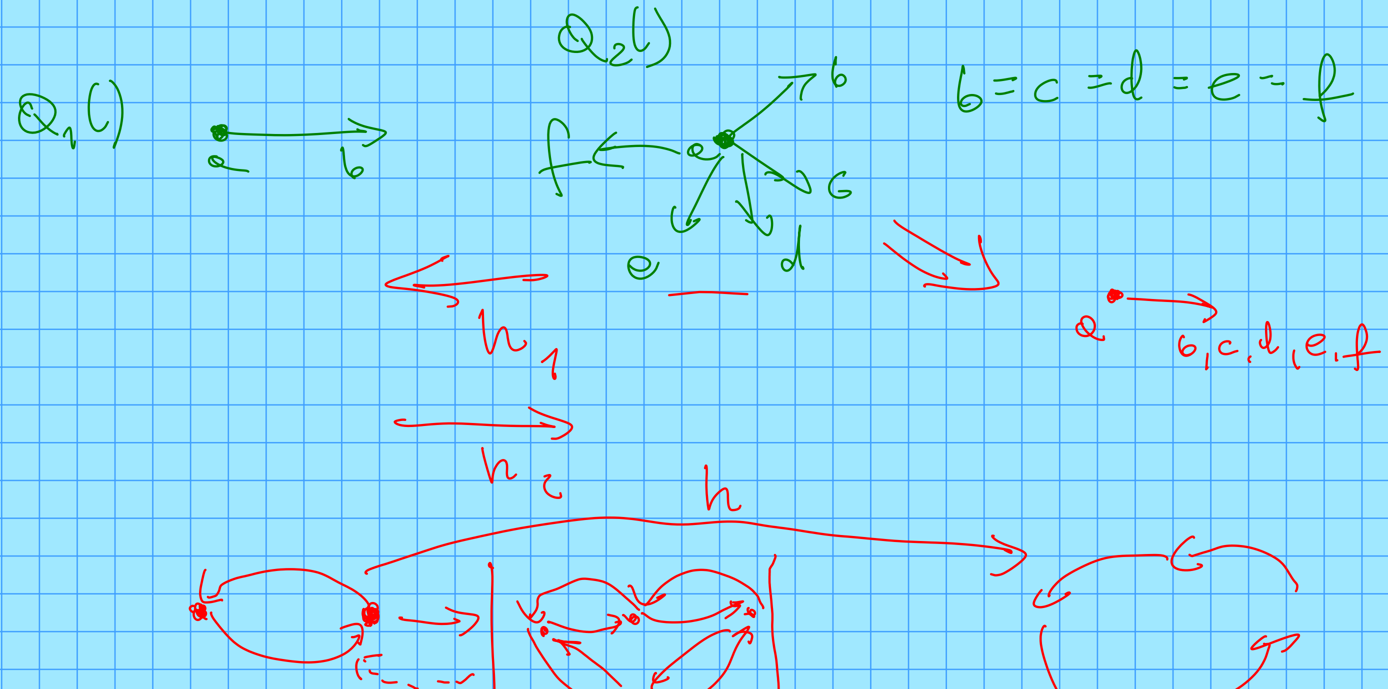
domknięcie przechodnie E $= \bigcup_{n \geq 1} T^n$

- $T^0 \subseteq T^1 \subseteq T^2 \subseteq \dots \subseteq T^n \subseteq \dots$

$T^i \subseteq adom(D) \times adom(D)$

Najdalej po $m \leq |adom(D)|^2$ krokach nasz algorytm się zatrzyma.

Data complexity ewaluacji zap. w datalogu jest wielomianowa.



$E(x,y), E(y,x)$