Matematické základy formální pojmové analýzy

(materiál k předmětu psychosémantika)

M. Filip

Formální pojmová analýza (angl. formal concept analysis, něm. formale Begriffsanalyse; dále jen FPA) je matematická metoda z oboru algebry, kterou vytvořil R. Wille v osmdesátých letech (Wille, 1996). Výchozí zobrazení dat pro FPA je tzv. formální kontext **K**, který obsahuje objekty z množiny O a atributy z množiny A. Vztahy mezi objekty a atributy jsou charakterizovány binární relací R, která určuje na základě vztahu oRa (také psáno $(o,a) \in R$), kdy má daný objekt o z O atribut a z A. Pro obecný popis formálního kontextu se používá výraz $\mathrm{K} = (O, A, I)$ [1]. Takto vymezený formální kontext je dobře zobrazitelný tabulkou, ve které jsou řádky obsazeny objekty, sloupce atributy a incidenční data vyjadřují relaci R. Tab. 1 ukazuje příklad formálního kontextu.

Tab. 1: Příklad formálního kontextu

		a Kouří dýmku	**b** Hodně jí	**c** Mluví nahlas	**d** Trpí nespavostí	**e** Nesnáší alkohol
1	Otec	0	1	1	0	0
2	Matka	0	0	0	1	1
3	Syn	0	1	0	0	0
4	Dcera	0	0	0	1	1
5	Děda	1	0	1	0	0

Pokud budeme uvádět pouze numerická a abecední označení objektů a atributů, má formální kontext zobrazený v této tabulce množinu objektů $O = \{1,2,...,5\}$ a množinu atributů $A = \{a, b, ...e\}$. Relace I je vyjádřena množinou $I = \{(1,b),(1,c),(2,d),(2,e),(3,b),(4,e),(5,a),(5,c)\}$. Pro vyjádření vztahů ve formálním kontextu použijeme operátor derivace, který definujeme pro množinu $X \subseteq O$ (Wille, 1996, s.18):

$$X' = \{a \in A \mid oIa\} \text{pro všechna } o \in X \qquad (1)$$

(množina společných atributů objektů z X).

Tomu odpovídá definice pro množinu $Y \subseteq A$:

[1] V citovaném německém originálu se objekty označují písmenem G (Gegendstände) a atributy písmenem M (Merkmale). Označení O a A přebírám z anglické literatury o FPA, protože je shodné i s českými termíny objekt a atribut.

$Y' = \{o \in O \mid oIa\}$ pro všechna $a \in Y$ (2)

(množina objektů, které mají společné atributy z Y).

Definovaný operátor se dá aplikovat vícekrát na tutéž množinu. Přitom platí: $X \subseteq X''$, $Y \subseteq Y''$ (další vztahy a jejich důkazy jsou obsaženy v citované knize).

Nyní definujeme *formální pojem*, který je ústřední jednotkou v FPA (Wille, 1996, s. 18):

„Formální pojem kontextu (O, A, I) je pár (X, Y), kdy $X \subseteq O, Y \subseteq A, X' = Y$ a $Y' = X$. X nazýváme *rozsahem-extenzí* a Y *obsahem-intenzí* pojmu (X, Y). $\mathrm{P}(O, A, I)$ označuje množinu všech pojmů kontextu (O, A, I)."

Formální pojem v našem konkrétním kontextu je např. pojem $(\{2,4\}, \{d, e\})$, protože všechny společné atributy objektů 2 a 4 jsou atributy *d, e* a naopak. Množina $\mathrm{P}(O, A, I)$ tohoto kontextu obsahuje celkem sedm formálních pojmů:

P1. $(\{1,2,3,4\}, \{\ \})$

P2. $(\{2,4\}, \{d, e\})$

P3. $(\{1,5\}, \{c\})$

P4. $(\{1,3\}, \{b\})$

P5. $(\{1\}, \{b, c\})$

P6. $(\{5\}, \{a, c\})$

P7. $(\{\ \}, \{a, b, c, d\})$.

FPA dále určuje hierarchické vztahy mezi formálními pojmy (Wille, 1996, s. 20):

„(X_1, Y_1) a (X_2, Y_2) jsou pojmy formálního kontextu. Pojem (X_1, Y_1) nazýváme podřazeným pojmu (X_2, Y_2) v případě, že $X_1 \subseteq X_2$ (stejně platné je přitom i kritérium $Y_2 \subseteq Y_1$). (X_2, Y_2) je potom nadřazený pojmu (X_1, Y_1) a píšeme $(X_1, Y_1) \le (X_2, Y_2)$. Relace $\le$ je nazývána jako hierarchické uspořádání pojmů. Takto uspořádanou množinu všech pojmů kontextu (O, A, I) označíme $\underline{\mathrm{P}}(O, A, I)$ a nazveme jako svaz pojmů (angl. concept lattice) kontextu (O, A, I)."

Podřazený pojem má tedy stejný nebo menší rozsah a stejný nebo širší obsah než jeho nadřazený pojem. Tato vlastnost odpovídá základní intuici tradiční sémantiky, že se vzrůstající konkrétností má pojem víc vlastností (je specifičtější) a zužuje se jeho rozsah. Ve směru k větší abstraktnosti platí pochopitelně opačná zákonitost.[2] Pro definování vlastností svazu pojmů slouží množinové operátory průniku a sjednocení. Pro jakoukoliv množinu formálních pojmů potom platí (Wille, 1996, s. 20):

„Svaz pojmů $\underline{P}(O, A, I)$ je úplný svaz, když pro infimum platí:

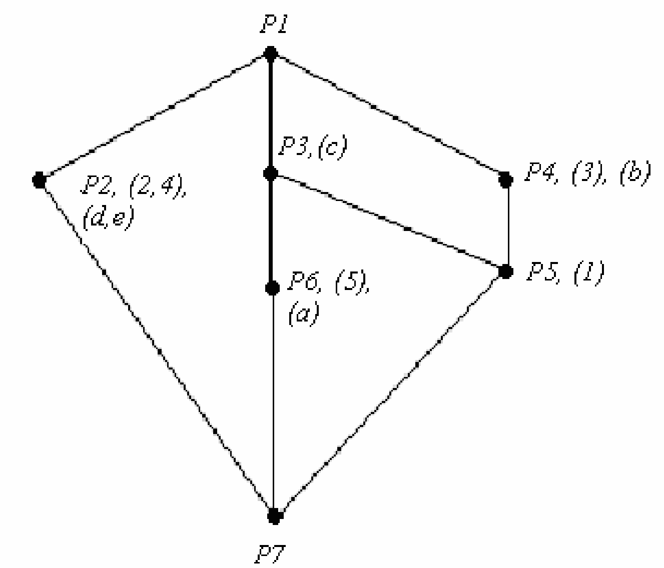
$$\bigwedge_{t\in T}(X_t, Y_t) = \left(\bigcap_{t\in T} X_t, \left(\bigcup_{t\in T} Y_t \right)'' \right) \tag{3}$$

a pro supremum platí:

$$\bigvee_{t\in T}(X_t, Y_t) = \left(\left(\bigcup_{t\in T} X_t \right)'', \bigcap_{t\in T} Y_t \right).\text{“} \tag{4}$$

Infimum, tj. podřazený pojem s nejširším obsahem, má vždy nulový rozsah a obsah daný jako sjednocení obsahů všech pojmů nadřazených. V našem příkladě je to pojem $(\{\ \}, \{a, b, c, d\})$. Supremum, tj. nadřazený pojem s největším rozsahem, má naopak vždy nulový obsah a rozsah rovný sjednocení rozsahů všech podřazených pojmů. V příkladu tedy představuje supremum pojem $(\{1,2,3,4\}, \{\ \})$.

Výsledky FPA se dají snadno zobrazit v grafu, kde uzly znamenají jednotlivé pojmy a spojnice mezi nimi relace $\leq$, jak to ukazuje obr. 1.



Obr. 1: Příklad grafické znázornění výsledků FPA (vysvětlení v textu)

[2] Podle moderní intenzionální sémantiky je však tato představa, založená na množinově-teoretickém přístupu, neudržitelná – viz přednáška o atomistické sémantice.

Indexy s *P* označují, který formální pojem kontextu výchozího příkladu daný uzel znázorňuje. Čísla charakterizují umístění objektů a malá písmena zase pozici atributů ve struktuře. Grafická struktura má vlastnosti korespondující s předchozími definicemi. *Jakýkoliv objekt na určitém místě struktury má atributy, které nese pojem, jehož je součástí, plus atributy všech nadřazených pojmů (tj. pojmů spojených vzestupně).* Např. objekt 5 má atribut *a* (zobrazený přímo u formálního pojmu P6) plus atribut *c* (zobrazený u nadřazeného formálního pojmu P5), což odpovídá výchozímu formálnímu kontextu. Naproti tomu *libovolný atribut obsazuje všechny ty objekty, které jsou ve struktuře na stejné úrovni a níže.* Např. atribut *c* obsazuje objekt 1 (zobrazený u formálního pojmu P5 podřazeného pojmu P3) a objekt 5 (zobrazený u formálního pojmu P6 podřazeného pojmu P5), což je rovněž v souladu s výchozím kontextem.

FPA se stává stále rozšířenější metodou v různých oblastech. Z psychologických disciplín našla dosud největší uplatnění v kognitivní psychologii (modelování mentálních reprezentací sémantických struktur), a v psychologii organizace a řízení (modelování struktur znalostí/vědění v rámci organizace). Potenciálně je použitelná i v kvalitativní metodologii jako nástroj pro vizualizaci struktur, které jsou v datech „skryty“.

Literatura:
Wille, R. (1996). *Formale Begriffsanalyse - Mathematische Grundlagen.* Berlin: Springer.

webové odkazy:

http://www.upriss.org.uk/fca/fca.html (základní stránka FPA)

http://www.mathematik.tu-darmstadt.de/ags/ag1/fag1/Software/software_en.html
(zde je volně ke stažení jednoduchý software pro FPA)

http://wundt.kfunigraz.ac.at/
(oddělení kognitivní psychologie v Grazu, která aplikuje FPA a příbuzné metody)