Analýza internetových vyhledávačů metodami formální konceptuální analýzy

Analysis of internet search engines using formal concept analysis method

Bc. Petr Fišnar

Diplomová práce
2012

Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky
akademický rok: 2011/2012

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Bc. Petr FIŠNAR**
Osobní číslo:	**A09693**
Studijní program:	**N 3902 Inženýrská informatika**
Studijní obor:	**Počítačové a komunikační systémy**
Téma práce:	**Analýza internetových vyhledávačů metodami formální konceptuální analýzy**

Zásady pro vypracování:

1. **V teoretické části zpracujte základní pojmy a tvrzení teorie uspořádaných množin a teorie svazů.**
2. **Zejména uveďte vlastnosti uzávěrových operátorů a větu o pevném bodě v úplných svazech. Tvrzení uvádějte bez důkazů, jen s odkazem na odbornou literaturu.**
3. **V praktické části zpracujte základy formální konceptuální analýzy a formulujte základní reprezentační větu pomocí Galoisových konexí. Uveďte konkrétní příklad kontextu a jeho konceptuálního svazu (z dané oblasti).**
4. **Uveďte přehled používaných internetových vyhledávačů a popište základní principy jejich činnosti.**
5. **Metodami formální konceptuální analýzy proveďte rozbor používaných internetových vyhledávačů a poskytovaných služeb z hlediska vytěžování informací.**

Rozsah diplomové práce:

Rozsah příloh:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

1. **WILLE, R., GANTER, B. Formal Concept Analysis – Mathematical Foundations. 1st ed. Springer, 1998. 284 s. ISBN 3-540-62771-5.**
2. **BĚLOHLÁVEK, Radim. Konceptuální svazy a formální konceptuální analýza [online]. [cit. 2009-05-15].**
3. **JAKUB, A. Teoretické informační systémy a vytěžování znalostí z databází [online]. [cit. 2009-05-15]. Dostupný z WWW: [http://www.users.fsid.cvut.cz/jura/pis/materialy/prednasky/TIVZFINI.doc].**
4. **DUCROU, J., EKLUND, P. SearchSleuh: The Conceptual Neighbourhood of an Web Query [online]. [cit. 2009-05-15]. Dostupný z WWW: [http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-331/Ducrou.pdf].**
5. **KOESTER, Bjoern. Conceptual Knowledge Processing with Google [online]. [cit. 2009-05-15]. Dostupný z WWW: [http://www.ke.informatik.tu-darmstadt.de/events/FGML-05/papers/paper11.pdf].**
6. **PRISS, Uta. Formal Concept Analysis in Information Science [online]. [cit. 2009-05-15]. Dostupný z WWW: [http://www.upriss.org.uk/papers/arist.pdf].**

Vedoucí diplomové práce:	**RNDr. Jiří Klimeš, CSc.** Ústav matematiky
Datum zadání diplomové práce:	**24. února 2012**
Termín odevzdání diplomové práce:	**28. května 2012**

Ve Zlíně dne 24. února 2012

L.S.

prof. Ing. Vladimír Vašek, CSc.
děkan

prof. Ing. Karel Vlček, CSc.
ředitel ústavu

ABSTRAKT

Diplomová práce „*Analýza internetových vyhledávačů metodami formální konceptuální analýzy*“ je zaměřena na problematiku internetových vyhledávačů, jejich analýzu z hlediska zaměření, principu činnosti nebo metod používaných při vyhledávání a řazení informací. Práce je rozdělena do dvou částí. V úvodním teoretickém přehledu jsou definovány základní termíny z odborné literatury týkající se zejména teorie svazů a jsou zde také vymezeny pojmy jako uzávěrový operátor nebo věta o pevném bodě. Praktická část práce představuje matematické základy formální konceptuální analýzy a způsoby její grafické interpretace a zabývá se především rozborem internetových vyhledávačů právě pomocí této metody.

Klíčová slova:

Teorie svazů, formální konceptuální analýza, atributy, objekty, formální kontext, formální koncept, konceptuální svaz, internetové vyhledávače.

ABSTRACT

The submitted thesis “*Analysis of internet search engines using formal concept analysis method* is focused on elaboration of internet search engines, their analysis in terms of focus, the principle of activity or methods used to search and sort information. The paper is divided in two parts. In the introductory overview, basic terms of the professional literature concerning to lattice theory and concepts like closure operator or fixed point theorem are defined. The practical part of the submitted thesis presents the mathematical foundations of formal concept analysis and ways of its graphical interpretation and deals mainly with the analysis of Internet search engines just by using this method.

Keywords:

Lattice theory, formal concept analysis, attributes, objects, formal context, formal concept, concept lattice, internet search engines.

Poděkování, motto

Chtěl bych touto formou poděkovat panu RNDr. Jiřímu Klimešovi, CSc. za vedení diplomové práce, za poskytování odborných rad, korektur a doporučení. Poděkování patří rovněž mojí přítelkyni a mé rodině za jejich podporu při studiu.

Prohlašuji, že

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji,

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně ……………………….

podpis diplomanta

OBSAH

ÚVOD

Nejrozšířenější techniku vyhledávání informací představuje v současné době bezesporu používání tzv. internetových vyhledávačů. Jde o populární službu umožňující uživateli najít takové webové stránky, které obsahují požadované informace o hledaných výrazech. Fungování internetových vyhledávačů je dnes již zřejmé téměř každému člověku – nejprve je nutné do rozhraní vyhledávače zadat klíčová slova charakterizující hledanou informaci a vyhledávač obratem na základě své databáze vypisuje seznam odkazů na stránky, které hledané informace obsahují (text, obrázky nebo jiné typy multimediálních informací).

Cílem internetových vyhledávačů je tedy přiblížit výsledek vyhledávání co nejblíže zadaným kritériím a poskytnout tak uživateli při odpovědi na dotaz co nejrelevantnější informace. K tomu je ovšem zapotřebí provést nejrůznějšími způsoby hodnocení obsahu a důležitosti webových stránek, které mají vyhledávače ve své databázi.

Trh vyhledávačů představuje velmi náročné ekonomické prostředí, které jednotlivé firmy neustále nutí do nejrůznějších technologických a marketingových inovací, jen aby si udržely svou pozici na trhu, případně aby svůj tržní podíl ještě rozšířily.

Současným trendem co nejvěrnějšího popisu reálných jevů je interdisciplinární přístup, který s sebou přináší velké množství analyzovaných proměnných, které je třeba nějakým způsobem statisticky zpracovat. Nabízí se široké spektrum metod, přičemž jednou z nich je i metoda formální konceptuální analýzy. Jedná se o moderní metodu analýzy dat založenou na myšlence seskupování zkoumaných objektů podle jim společných vlastností, čehož lze úspěšně využít při tvorbě výzkumných hypotéz v mnoha přírodních vědách. Tato metoda má komplexní matematický základ, který navazuje na práci německého matematika Rudolfa Willeho z 80. let minulého století.

Tato metoda umožňuje navrhnout na základě dat, získaných například pomocí dotazníků nebo přístrojově naměřených, tzv. konceptuální svaz – datovou strukturu, která je uživateli snadno zobrazitelná, a která expertovi usnadňuje hledání zajímavých souvislostí v datech. Expert pak může dále konceptuálním svazem procházet a pozorovat různé závislosti mezi atributy. Konceptuální svaz mu nabízí první, avšak detailně strukturovaný pohled na data. Na základě tohoto počátečního vhledu může provádět podrobnější analýzy, při kterých lze již využít běžné statistické metody.

Cílem této diplomové práce je zpracovat základy formální konceptuální analýzy a formulovat základní reprezentační větu pomocí Galoisových konexí, uvést přehled jednotlivých internetových vyhledávačů a popsat základní principy jejich činnosti a provést rozbor internetových vyhledávačů a poskytovaných služeb právě metodami formální konceptuální analýzy.

Diplomová práce je rozdělena do dvou částí. V literárním přehledu jsou definovány základní termíny z odborné literatury, které řeší problematiku teorie svazů. Podrobněji jsou zde vymezeny jednotlivé pojmy jako například polosvazy, svazy, úplné svazy a uzávěrové operátory a věta o pevném bodě.

Praktická část diplomové práce představuje matematické základy formální konceptuální analýzy a způsoby její grafické interpretace a zabývá se rozborem internetových vyhledávačů právě pomocí této metody. Vlastní práce byla sestavena především na základě teoretických poznatků z odborné literatury. Obsahuje základní definice o formální konceptuální analýze a popisuje jednotlivé internetové vyhledávače. Poslední kapitola pak na základě zjištěných teoretických znalostí aplikuje metodu formální konceptuální analýzy na konkrétním příkladu.

I. TEORETICKÁ ČÁST

1 ZÁKLADY TEORIE SVAZŮ

Matematický pojem teorie svazů patří do oblasti algebry, kde mezi uspořádanými množinami omezuje ty, která zachovávají *supremum* $\sup(x, y)$ a *infimum* $\inf(x, y)$. Supremum je zaváděno jako alternativa k pojmu největší prvek, oproti největšímu prvku je však dohledatelné u více množin – například omezené otevřené intervaly reálných čísel nemají největší prvek, ale mají supremum. Opakem suprema je infimum [5].

V první části práce budou vysvětleny základní pojmy teorie svazů. Připomeneme si některé pojmy, jako je například svaz, který je potřebný pro orientaci v dalším textu.

Svaz můžeme definovat buď jako algebraickou strukturu nebo také jako částečně uspořádanou množinu, která splňuje určité vlastnosti. Podíváme se na svazy oběma směry a ukážeme si spojitost obou definicí [5].

Dále si řekneme, co jsou svazy modulární a distributivní a definujeme vlastnosti, ve kterých se jednotlivé druhy svazu odlišují. V práci bude uvedeno několik příkladů těchto objektů a budou uvedeny jejich základní vlastnosti.

1.1 Definice svazů

1.1.1 Částečně uspořádaná množina

Částečně uspořádaná množina formalizuje uspořádání (určení pořadí některých prvků) na množině. Skládá se z množiny a binární relace popisující uspořádání jednotlivých dvojic prvků. Definujeme ji formálně. Částečné uspořádání je binární relace $\leq$ na množině G, která je

→ *reflexivní* – pro každé $x, y, z \in G$ platí $a \leq a$,

→ *tranzitivní* – $x \leq y$ a zároveň $y \leq z$, pak platí $x \leq z$,

→ *antisymetrická* – pokud $x \leq y$ a $y \leq x$, pak platí $x = y$.

Uspořádanou množinou uvažujeme dvojici $(G, \leq)$, tedy množinu a částečné uspořádání na ní definované.

Příklad 1.1.1.1. Množina podmnožin dané množiny s relací inkluze.

Příklad 1.1.1.2. Množina přirozených čísel s relací dělitelnosti.

Příklad 1.1.1.3. Množina reálných čísel s relací menší nebo rovno $\leq$ [5].

1.1.2 Algebraická struktura

V úvodu bylo uvedeno, že svaz je možné definovat jako algebraickou strukturu, přesněji jako množinu s dvěma binárními operacemi. Jde o operace $\wedge$ (průsek) a operaci $\wedge$ (spojení).

Předpokládejme, že struktura $(G, \wedge, \vee)$ je svazem, pak pro všechny prvky $x, y, z \in G$ platí

→ *komutativita* – $x \wedge y = y \wedge x, x \vee y = y \vee x$,

→ *asociativita* – $(x \wedge y) \vee z = x \wedge (y \vee z), (x \vee y) \wedge z = x \vee (y \wedge z)$,

→ *absorpce* – $x \vee (y \wedge x) = x, x \wedge (y \vee x) = x$,

z toho plyne

→ *idempotence* – $x \wedge x = x, x \vee x = x$ [5].

1.1.3 Spojitost obou definicí

Svaz vyjádřený uspořádanou množinou lze snadno vyjádřit algebraickou strukturou. Relaci částečného uspořádání $\leq$ vyjádříme pomocí operací $\wedge$ a $\vee$. Mějme svaz $(G, \wedge, \vee)$. Částečné uspořádání $\leq$ na množině G, můžeme definovat takto

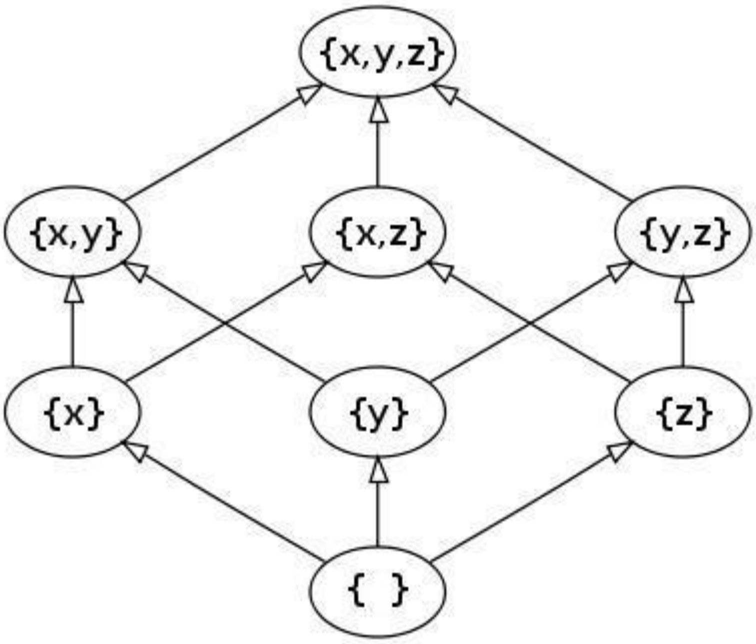
$x \leq_G y$ tehdy a jen tehdy, pokud $x \wedge y = x$ resp. $x \vee y = y$.

Příklad 1.1.3.1. Mějme svaz $(G, \leq)$, kde $G = 2^{\{x,y,z\}}$ a operace částečného uspořádání $\leq$ odpovídá množinové inkluzi $\subseteq$. Dále svaz $(H, \wedge, \vee)$, kde operace průseku a spojení odpovídají množinovému průniku $\cap$ a sjednocení $\cup$. Potom je H ekvivalentní s G. Dá se ukázat, že vezmeme-li dva libovolné prvky $x, y, z \in G$, pak platí, že jsou-li v relaci $x \subseteq y$, pak také platí, že $x \wedge y = x$ a $x \vee y = y$. Například zvolme $x = \{k\}$, $y = \{k, l\}$. Prvky x, y jsou v relaci $\leq$, (protože $x \subseteq y$) a zároveň také platí, že $\{x\} \cup \{x, y\} = \{x, y\}$ a $\{x\} \cap \{x, y\} = \{x\}$ [5].

1.2 Hasseův diagram

Hasseův diagram se používá pro znázornění relace uspořádání na množině. Jeho název je odvozen od německého matematika Helmuta Hassea.

Jde o orientovaný graf, podobný uzlovému grafu, jehož vrcholy reprezentují jednotlivé prvky částečně uspořádané množiny a hrany znázorňují relaci pokrytí příslušnou danému uspořádání. Hasseův diagram je zobrazení, kde častěji větší prvek vzhledem k uspořádání je umístěn výše než prvek menší. Proto se někdy šipky u hran vynechávají. Umístění vrcholů je vhodné volit tak, aby pokud možno nedocházelo ke křížení hran. Hasseův diagram na obrázku níže znázorňuje množinu 2^X množiny $X = \{x, y, z\}$ [5].



Obr. 1. Hasseův diagram

1.3 Svaz

Definice: Uspořádaná množina, v níž ke každým dvěma prvkům existuje supremum i infimum, se nazývá svaz.

Uspořádaná množina se dvěma binárními operacemi, které jsou asociativní, komutativní, idempotentní a splňují zákony absorpce, se nazývá *svaz*.

Věta 1.3.1. Nechť $(G, \leq)$ je uspořádaná množina, kde pro každé $x, y \in G$ existuje $\sup(x, y) = x \vee y$ a $\inf(x, y) = x \wedge y$. Pak $(G, \vee, \wedge)$ je svaz, $(G, \vee)$ a $(G, \wedge)$ jsou polosvazy, kde obě operace jsou spolu svázány tzv. *absorpčními zákony*, kde pro každé prvky $x, y \in G$ platí

$$x \vee (x \wedge y) = \sup\ (x, \inf\ (x, y)) = x,$$

$$x \wedge (x \vee y) = \inf\ (x, \sup\ (x, y)) = x.$$

Kromě toho pro každé prvky $x, y \in G$ platí

$$x \wedge y = x \Leftrightarrow x \leq y \Leftrightarrow x \vee y = y.$$

Věta 1.3.2. Nechť $(G, \vee, \wedge)$ je množina se dvěma idempotentními, asociativními a komutativními operacemi, které jsou spolu svázány absorpčními zákony. Pak

→ pro každé prvky $x, y \in G$ platí

$$x \wedge y = x \Leftrightarrow x \vee y = y,$$

→ definujeme-li na G relaci $\leq$ takto, pak pro libovolné prvky $x, y \in G$ klademe

$$x \leq y \Leftrightarrow x \vee y = y,$$

pak je $\leq$ uspořádání na G takové, že $(G, \leq)$ je svaz, v němž pro libovolné prvky $x, y \in G$ je prvek $x \vee y$ jejich supremum a prvek $x \wedge y$ jejich infimum.

Z uvedených vět vyplývá, že svazy jsou totéž co algebraické struktury $(G, \vee, \wedge)$ se dvěma idempotentními, asociativními a komutativními operacemi, svázanými spolu absorpčními zákony. Proto tyto struktury $(G, \vee, \wedge)$ budeme nazývat *svazy*.

Princip duality: Je-li $(G, \vee, \wedge)$ svaz, pak i $(G, \wedge, \vee)$ je svaz. Obecně, jestliže v nějakém platném tvrzení o svazech systematicky zaměníme supremum za infimum, $\vee$ za $\wedge$, $\leq$ za $\geq$ dostaneme opět platné tvrzení o svazech.

Protože není nutné zdůrazňovat, zda máme na mysli svaz jako uspořádanou množinu nebo jako algebraickou strukturu se dvěma operacemi, nebudeme v dalším textu, nebude-li to z určitého důvodu vhodné nebo dokonce nevyhnutelné, uspořádání či operace vyznačovat. Budeme tedy místo o svazu $(G, \leq)$ či svazu $(G, \vee, \wedge)$ jednoduše psát o svazu G.

Věta 1.3.3. V libovolném svazu G pro každou trojici prvku $x, y, z \in G$ platí tzv. *distributivní nerovnosti*

$$(x \vee y) \wedge (x \vee z) \geq x \vee (y \wedge z),$$

$$(x \wedge y) \vee (x \wedge z) \leq x \wedge (y \vee z).$$

Je-li navíc $z \leq x$, platí tzv. *modulární nerovnost*

$$(x \wedge y) \vee z \leq x \wedge (y \vee z).$$

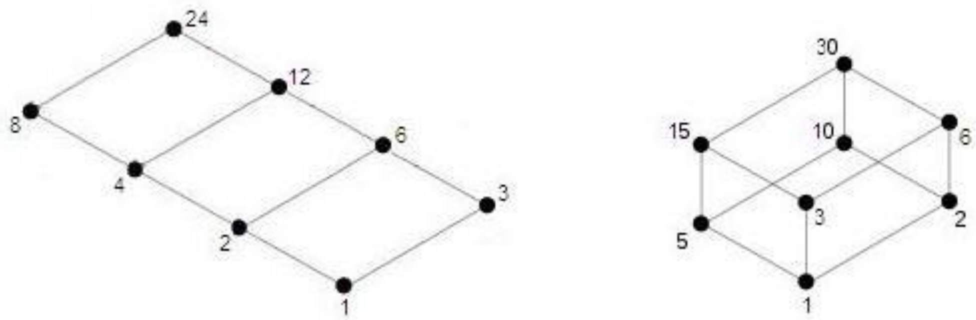
Věta 1.3.4. Nechť G je svaz, $n \in \mathbb{N}$. Pro libovolné prvky $x_1, \dots, x_n \in G$ platí, že $x_1 \vee \dots \vee x_n$ je supremum množiny $\{x_1, \dots, x_n\}$ a $x_1 \wedge \dots \wedge x_n$ je infimum množiny $\{x_1, \dots, x_n\}$.

Příklad 1.3.1. Každý řetězec (neboli lineárně uspořádaná množina, tj. uspořádaná množina, v níž jsou každé dva prvky srovnatelné) je svaz.

Příklad 1.3.2. Pro libovolnou množinu X je $(2^X, \subseteq)$ svaz.

Příklad 1.3.3. Nechť $\mathbb{N}$ je množina přirozených čísel, $x \vee y$ je NSN (x, y) → nejmenší společný násobek čísel x, y, $x \wedge y$ je NSD (x, y) → největší společný dělitel čísel x, y, pak $(\mathbb{N}, \vee, \wedge)$ je svaz.

Příklad 1.3.4. Nechť n je přirozené číslo, $P(n)$ je množina všech dělitelů čísla n. Pak $(P(n), NSN, NSD)$ je svaz. Pro $n = 24$ resp. $n = 30$ je tento svaz znázorněn diagramem na následujícím obrázku [1], [2], [6].



Obr. 2. Znázornění svazu pro $n = 24$, resp. $n = 30$

1.4 Polosvaz

Představme si grupoid $(G, \wedge, \vee)$, kde pro každé prvky $x, y, z \in G$ platí

→ *komutativita* – $x \wedge y = y \wedge x$, $x \vee y = y \vee x$,

→ *asociativita* – $(x \wedge y) \vee z = x \wedge (y \vee z)$, $(x \vee y) \wedge z = x \vee (y \wedge z)$,

→ *absorpce* – $x \vee (y \wedge x) = x, x \wedge (y \vee x) = x$,

z toho plyne

→ *idempotence* – $x \wedge x = x, x \vee x = x$.

Komutativní idempotentní pologrupu nazýváme *polosvazem*.

Věta 1.4.1. Nechť $(G,\wedge,\vee)$ je komutativní pologrupa. Pak množina všech idempotentních prvků tvoří podgrupoid pologrupy $(G,\wedge,\vee)$, který je polosvazem.

Věta 1.4.2. Nechť $(G,\cdot)$ je polosvaz. Potom relace $\leq$ definovaná na G předpisem

$$x \leq y \Leftrightarrow x \cdot y = y,$$

je uspořádání, přičemž v uspořádané množině $(G, \leq)$ existuje pro každé $x, y \in G$ $\sup\ (x, y)$ a platí $\sup\ (x, y) = x \cdot y$.

Věta 1.4.3. Nechť $(G, \leq)$ je uspořádaná množina a pro každé $x, y \in G$ existuje $\sup\ (x, y)$. Označme $x \cdot y = \sup\ (x, y)$. Pak $(G, \leq)$ je polosvaz.

$$x \leq y \Leftrightarrow x \vee y = y$$

Z vět uvedených výše plyne:

Polosvazy jsou totéž co uspořádané množiny, kde ke každým dvěma prvkům existuje supremum.

Zaměníme-li uspořádání za tzv. *duální*, tj. $\geq$, pak dle principu duality, pojem suprema v duálním uspořádání je $\geq$, infimem v uspořádání $\leq$. Definujeme-li v polosvazu uspořádání předpisem

$$x \leq y \Leftrightarrow x \cdot y = x,$$

pak pro každé $x, y \in G$ existuje $\inf (x, y)$ a platí $\inf\ (x, y) = x \cdot y$. Obráceně, je-li $(G, \leq)$ uspořádaná množina, kde pro každé $x, y \in G$ existuje $\inf\ (x, y)$, pak $(G,\cdot)$, kde $x \cdot y = \inf (x, y)$ je polosvaz. Z principu duality plyne:

Polosvazy jsou totéž co uspořádané množiny, kde ke každým dvěma prvkům existuje infimum.

Příklad 1.4.1. Pro libovolnou množinu X budeme symbolem 2^X označovat množinu všech podmnožin množiny X. Pak $(2^X,\cap)$ a $(2^X,\cup)$ jsou polosvazy.

Příklad 1.4.2. Na množině všech přirozených čísel $\mathbb{N}$ zaveďme relaci dělitelnosti $x|y$ právě, když x dělí y. Pak | je uspořádáním na $\mathbb{N}$ a každé $x, y \in \mathbb{N}$ je

→ $\sup(x, y) \rightarrow \text{NSN } (x, y)$,

→ $\inf(x, y) \rightarrow \text{NSD } (x, y)$.

Tedy $(\mathbb{N}, \text{NSN})$, $(\mathbb{N}, \text{NSD})$ jsou polosvazy [1], [2], [6].

1.4.1 Podsvaz

Definice: Nechť $(G,\vee,\wedge)$ je svaz, nechť $\emptyset \neq H \subseteq G$. H se nazývá podsvaz svazu $(G,\vee,\wedge)$, jestliže $\forall x,\ y \in H$ platí $x \vee y \in H$, $x \wedge y \in H$.

Nechť $(H, \leq)$ je uspořádaná množina, $x, y \in H$ a platí $x \leq y$. Množina $[x, y] = \{n \in H; x \leq n \leq y\}$ se nazývá *interval* v H.

Věta 1.4.1.1. Nechť $(G,\vee,\wedge)$ je svaz, Hi pro $i \in I$ jeho podsvazy. Je-li $\cap \{H_{ij} i \in I\} \neq \emptyset$, pak je $\cap \{H_{ij} i \in I\}$ podsvazem svazu $(G,\vee,\wedge)$.

Věta 1.4.1.2. Nechť $(G,\vee,\wedge)$ je svaz. Pak platí

→ pro každý prvek $x \in G$ je $\{x\}$ podsvaz svazu $(G,\vee,\wedge)$,

→ každý interval svazu $(G,\vee,\wedge)$ je jeho podsvaz,

→ má-li G prvky 0 a 1, pak $G = [0,1]$.

Příklad 1.4.1.1. Každá jednoprvková podmnožina svazu je jeho podsvazem, prázdná množina je podsvazem libovolného svazu, každý svaz je svým podsvazem [1], [6].

1.4.2 Ideál a filtr

Definice: Nechť G je svaz, $H \subseteq G$ podmnožina. Řekneme, že H je ideál svazu G, jestliže je H podsvazem svazu G, který navíc splňuje podmínku, že pro každé $x \in H$ a každé $k \in G$ platí

$$k \leq x \Rightarrow k \in H.$$

Duálně řekneme, že H je filtr svazu G, jestliže je H podsvazem svazu G, který navíc splňuje podmínku, že pro každé $x \in H$ a každé $k \in G$ platí

$$k \geq x \Rightarrow k \in H.$$

Ideál svazu je tedy podsvaz, který s každým svým prvkem x obsahuje i všechny prvky svazu menší než x, *filtr* svazu je podsvaz, který s každým svým prvkem x obsahuje i všechny prvky svazu vetší než x.

Věta 1.4.2.1. Průnik libovolného neprázdného systému podsvazů (resp. ideálů, resp. filtrů) daného svazu je opět podsvaz (resp. ideál, resp. filtr) tohoto svazu.

Nechť G je svaz, $H \subseteq G$ podmnožina. Díky předchozí větě můžeme nyní definovat ideál $H \downarrow$ svazu G generovaný množinou H jako průnik všech ideálů tohoto svazu obsahujících množinu H. Duálně, filtr $H \uparrow$ svazu G generovaný množinou H je průnik všech filtrů tohoto svazu obsahujících množinu H. Je-li $H = \{x\}$, píšeme stručně $x \downarrow$ místo $\{x\} \downarrow$, resp. $x \uparrow$ místo $\{x\} \uparrow$, hovoříme o *hlavním ideálu*, resp. o *hlavním filtru*, generovaném prvkem x.

Ideál H ve svazu G se tedy nazývá hlavní, právě když existuje prvek $x \in G$ takový, že $H = (\leftarrow, x >$. Jestliže $0 \in G$, pak ideál $\{0\}$ se nazývá *nulový.*

Duálně pak filtr H ve svazu G se nazývá hlavní, právě když existuje prvek $x \in G$ takový, že $H =< x, \rightarrow)$. Jestliže $1 \in G$, pak filtr $\{1\}$ se nazývá *jednotkový*.

Je zřejmé, že podmnožina $H \subseteq G$ je ideálem svazu G, právě když $H \downarrow = H$, a je filtrem svazu G, právě když $H \uparrow = H$.

Věta 1.4.2.2. Nechť G je svaz, $H \subseteq G$ podmnožina. Pro ideál $H \downarrow$ generovaný množinou H platí

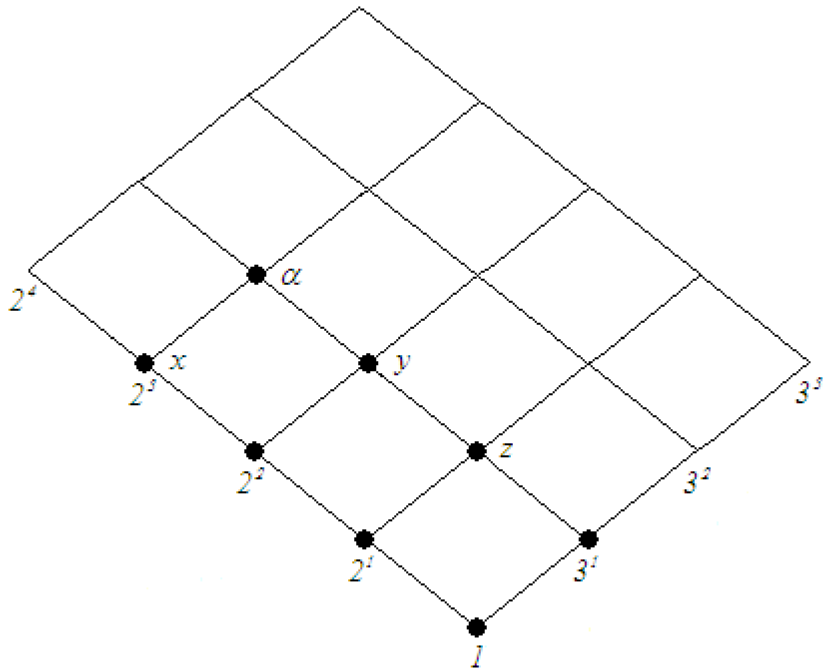
$$H \downarrow = \{k \in G; \exists n \in \mathbb{N}\ \exists x_1, \dots, x_n \in H: k \leq x_1 \vee \dots \vee x_n\}.$$

Duálně, pro filtr $H \uparrow$ generovaný množinou H platí

$$H \uparrow = \{k \in G; \exists n \in \mathbb{N}\ \exists x_1, \dots, x_n \in H: k \geq x_1 \wedge \dots \wedge x_n\}.$$

Příklad 1.4.2.1. Každý svaz je svým ideálem i filtrem. Prázdná množina je ideálem i filtrem libovolného svazu.

Příklad 1.4.2.2. Uvažujme svaz G všech celočíselných kladných dělitelů čísla $2^4 \cdot 3^3$, jehož Hasseův diagram je znázorněn na obrázku níže, pak ideál H generovaný množinou $\{a, b, c\}$, kde $a = 2^3$, $b = 2^2 \cdot 3^1$ a $c = 2^1 \cdot 3^1$ je hlavní ideál generovaný prvkem $d = a \vee b \vee c = 2^3 \cdot 3^1$. Tento ideál je znázorněn na obrázku níže plnými kroužky [3], [6].



Obr. 3. Hasseův diagram se znázorněným ideálem

1.4.3 Homomorfismus

Definice: Nechť $(G,\vee,\wedge)$ a $(H,\vee,\wedge)$ jsou svazy. Zobrazení $h : G \to H$ se nazývá *homomorfismus*, jestliže pro každé $x, y \in G$ platí

$$h(x \vee y) = h(x) \vee h(y),$$

$$h(x \wedge y) = h(x) \wedge h(y).$$

Homomorfismus je tedy zobrazení z jedné algebraické struktury do jiné stejného typu, které zachovává veškerou důležitou strukturu [7].

Homomorfismus, který je bijektivním zobrazením, nazýváme *izomorfismus*.

Věta 1.4.3.1 Budou-li G, H, I svazy, $f: G \to H$ a $g: H \to I$ homomorfismy (resp. izomorfismy). Pak složené zobrazení $f \cdot g$ je opět homomorfismus (izomorfismus) $G \to I$. Je-li $f: G \to H$ izomorfismus, je i $f^{-1}: H \to G$ izomorfismus. Identické zobrazení $id(x) = x$ je izomorfem svazu G. Je-li h homomorfismus svazu G do H, $x, y \in G$ a platí $x \leq y$ vzhledem k indukovanému uspořádání na G, pak $h(x) \leq h(y)$ vzhledem k indukovanému uspořádání na H.

Věta 1.4.3.2. Nechť G a H jsou svazy, $f: G \to H$ zobrazení

→ je-li f svazový homomorfismus, pak f je izotonní zobrazení a homomorfní obraz

$$f(G) = \{f(x); x \in G\}$$

je podsvaz svazu H.

→ zobrazení f je svazový izomorfismus, právě když f je izomorfismus uspořádaných množin.

Jestliže existuje izomorfismus svazu G na svaz H říkáme, že svazy G, H jsou izomorfní a označujeme je $G \cong H$.

Vzhledem k **Větě 1.4.3.1** platí, že jsou-li G, H, I svazy, pak

$$G \cong G,$$

$$G \cong H \Rightarrow H \cong G,$$

$$G \cong H \; a \; H \cong I \Rightarrow G \cong I.$$

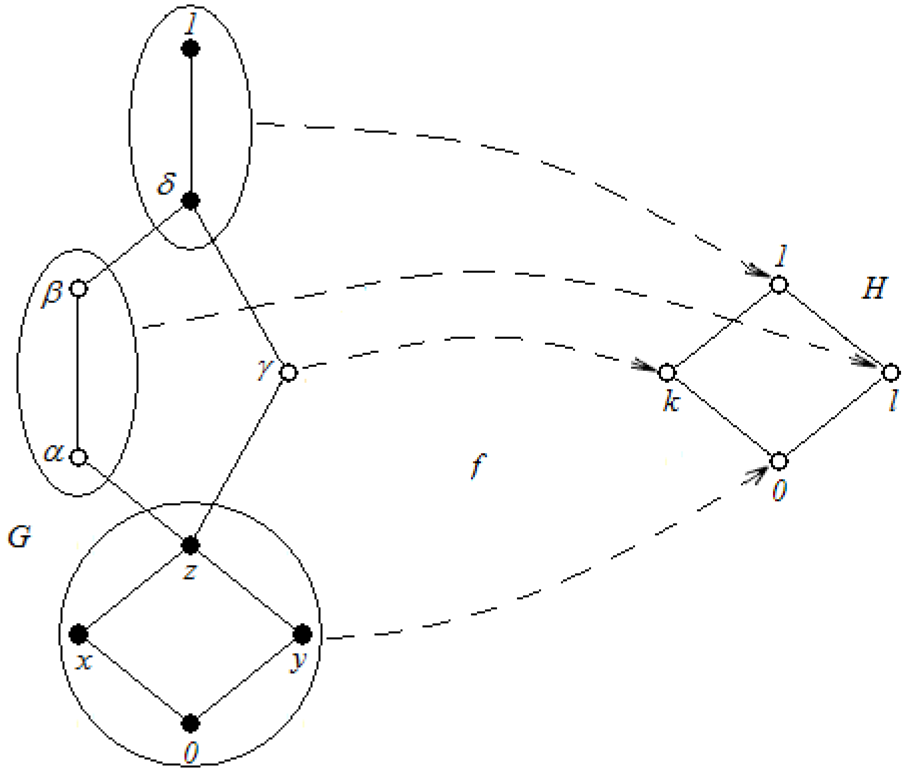
Tedy relace izomorfní je ekvivalencí na třídě všech svazů.

Příklad 1.4.3.1. Uvažujeme svazy G a H znázorněné na obrázku níže. Definujeme-li zobrazení $f: G \to H$ takto

$$f(0) = f(x) = f(y) = f(z) = 0,$$

$$f(\alpha) = f(\beta) = l, f(\gamma) = k, f(\delta) = f(1) = 1,$$

pak zobrazení f je svazový homomorfismus G na H [1], [2], [6].



Obr. 4. Znázornění svazového homomorfismus

1.4.4 Kongruence svazu

Definice: Nechť $(G,\vee,\wedge)$ je svaz a θ je ekvivalence na G. Pak θ se nazývá *kongruence svazu* G, platí-li

$$\forall x, y, z, \alpha \in G; (x, z) \in \theta, (y, \alpha) \in \theta \Rightarrow (x \vee y, z \vee \alpha) \in \theta, (x \wedge y, z \wedge \alpha) \in \theta.$$

Pro libovolný prvek $x \in G$ budeme třídu kongruence θ obsahující tento prvek označovat $[x]_\theta$, popř. $[x]$ (tj. $[x]_\theta = [x] = \{k \in G; (x, k) \in \theta\}$).

Jestliže $(G, \leq)$ je uspořádaná množina a $H \subseteq G$, pak H se nazývá *konvexní podmnožina* uspořádané množiny G, platí-li

$$\forall x, y \in H, k \in G; x \leq k \leq y \Rightarrow k \in H.$$

Jestliže G je navíc svaz a H je jeho podsvaz, který je současně konvexní podmnožinou v G, pak H se nazývá *konvexní podsvaz* svazu G.

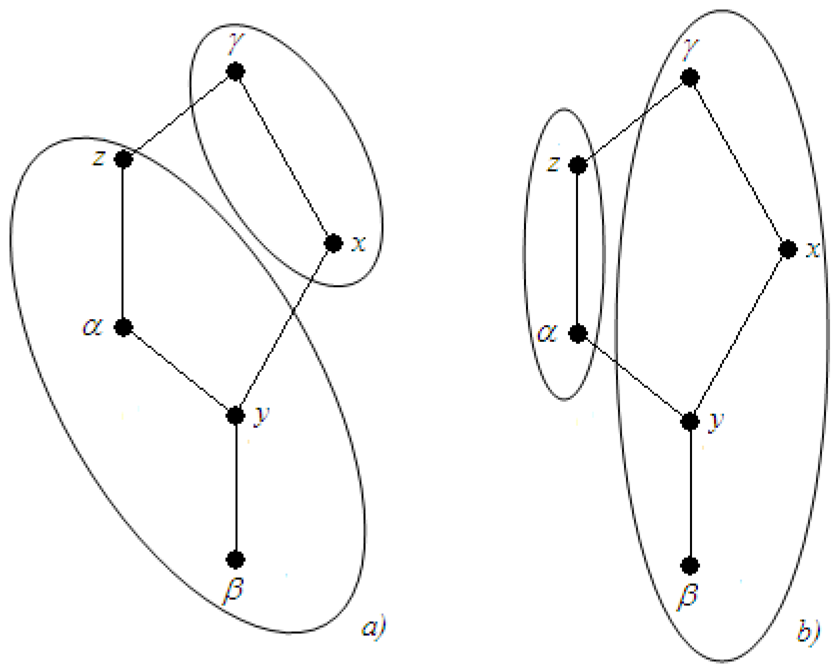
Jestliže x, y jsou prvky uspořádané množiny G takové, že $x \leq y$, pak (uzavřeným) intervalem mezi prvky x a y se rozumí množina

$$[x, y] = \{k \in G; x \leq k \leq y\}.$$

Platí tedy, že podmnožina $H \subseteq G$ je konvexní v G právě tehdy, když s každými svými dvěma prvky x a y takovými, že $x \leq y$ obsahuje i celý interval $[x, y]$.

Věta 1.4.4.1. Jestliže $(G,\vee,\wedge)$ je svaz, θ je kongruence svazu G a $x \in G$, pak třída $[x]_\theta$ je konvexním podsvazem svazu G [4].

Příklad 1.4.4.1. Uvažujeme svaz G, jehož Hasseův diagram je znázorněn na obrázku níže. Je zde na obrázku níže (a) vyznačen rozklad množiny G odpovídající kongruenci v G, zatímco na obrázku níže (b) je vyznačen rozklad odpovídající ekvivalenci E, která není kongruencí v G. Platí např. $a\ E\ b$, ale neplatí $a \vee c\ E\ b \vee c$, neboť prvky $a \vee c = f, b \vee c = c$ leží v různých blocích rozkladu G/E [3].



Obr. 5. Rozklad množiny G odpovídající kongruenci v G a) a ekvivalenci b)

1.4.5 Kartézský součin svazů

Tak jako lze v teorii grup vytvářet pomocí kartézského součinu z daných grup nové, můžeme analogicky postupovat i u svazů.

Věta 1.4.5.1. Nechť $(G;\vee,\wedge)$ a $(H;\vee,\wedge)$ jsou svazy, pak algebraická struktura s nosičem $G \times H$, jejíž operace $\vee$ a $\wedge$ jsou definovány takto

$$(x_1, x_2) \wedge (y_1, y_2) = (x_1 \wedge y_1, x_2 \wedge y_2),$$

$$(x_1, x_2) \vee (y_1, y_2) = (x_1 \vee y_1, x_2 \vee y_2),$$

je svaz.

Podstata důkazu **Věty 1.4.5.1.** spočívá v tom, že při provádění operací s dvojicemi $[x_1, x_2] \in G \times H$ se s prvními složkami dělají svazové operace z G a nezávisle na tom se s druhými složkami dělají svazové operace z H. Na základě **Věty 1.4.5.1.** lze vyslovit následující definici.

Definice: Nechť $(G;\vee,\wedge)$ a $(H;\vee,\wedge)$ jsou svazy, pak svaz $G \times H$, jehož operace $\vee$ a $\wedge$ jsou definovány formulemi ve **Větě 1.4.5.1**, se nazývá kartézský součin svazů G, H.

Jestliže bude potřeba zavést ve svazu $G \times H$ svazové uspořádání, pak to uděláme známým způsobem

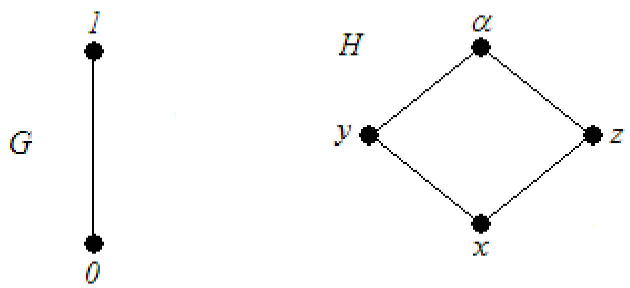
$$[x_1, x_2] \leq [y_1, y_2] \Leftrightarrow [x_1, x_2] \wedge [y_1, y_2] = [x_1, x_2],$$

což jinak formulováno znamená

$$[x_1, x_2] \leq [y_1, y_2] \Leftrightarrow x_1 \leq y_1 \wedge x_2 \leq y_2.$$

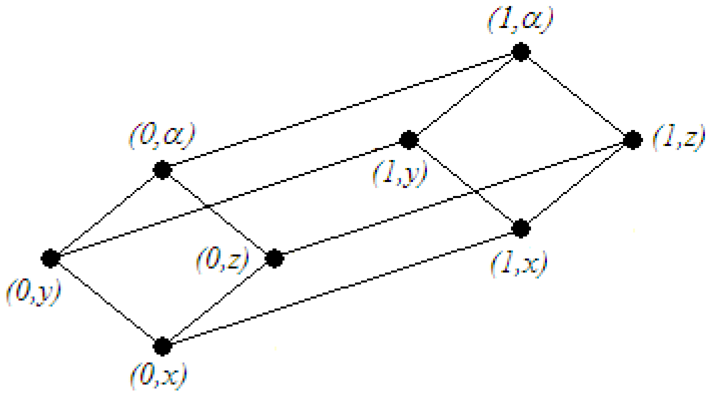
Protože operace kartézský součin je asociativní, lze ji rozšířit na libovolný konečný, popř. i nekonečný soubor svazů. Tuto operaci můžeme zavést i pro polosvazy. [3]

Příklad 1.4.5.1. Nechť $G = (\{0,1\}, \leq_G)$ a $H = (\{x, y, z, \alpha\}, \leq_H)$ jsou svazy určené diagramy na následujícím obrázku.



Obr. 6. Diagramy určených svazů G a H

Pak svaz I, který je přímým součinem G a H, má diagram znázorněný na obrázku níže [4].



Obr. 7. Diagram svazu $I(G \times H)$

1.5 Úplný svaz

Podle **Věty 1.3.1.** je svazem každá uspořádaná množina $(G, \leq)$, ve které existují sup (x, y) a inf (x, y) pro každé dva prvky $x, y \in G$. Indukcí lze snadno dokázat, že je-li $(G, \leq)$ svazem, pak existují supremum a infimum pro každou konečnou podmnožinu $H \subseteq G$. Nelze odtud však odvodit žádné tvrzení o supremum a infimum nekonečných podmnožin. Existují však svazy, ve kterých supremum a infimum existují i pro nekonečné podmnožiny. Zavedeme proto následující pojem.

Definice: Uspořádaná množina $(G, \leq)$ se nazývá úplný svaz, jestliže pro každou $H \subseteq G$ existuje supremum H a infimum H v G.

Každý úplný svaz G má nejmenší prvek (infimum množiny G ve svazu G) a největší prvek (supremum množiny G ve svazu G).

Je-li $H \subseteq G$, pak infimum množiny H ve svazu G je největší dolní závora množiny H ve svazu G. Dolní závora množiny H ve svazu G je prvek $k \in G$ takový, že pro každé $x \in H$ platí $k \leq x$. V případě $H = \emptyset$ je tato podmínka splněna pro každé $k \in G$, a tedy odtud plyne, že každý prvek svazu G je v G dolní závorou prázdné množiny. Proto infimem prázdné množiny ve svazu G je největší prvek svazu G. Duálně platí, že supremem prázdné množiny ve svazu G je nejmenší prvek svazu G.

Příklad 1.5.1. Platí, že každý úplný svaz je svazem a podle **Věty 1.3.1.** je každý neprázdný konečný svaz úplným svazem.

Příklad 1.5.2. Prázdný svaz není úplný, neboť pro jeho (jedinou) prázdnou podmnožinu neexistuje infimum ani supremum. Jinými slovy prázdný svaz nemá nejmenší prvek ani největší prvek, protože nemá žádný prvek.

Příklad 1.5.3. Pro libovolnou množinu X je $(2^X, \subseteq)$ úplný svaz.

Příklad 1.5.4. Nekonečný řetězec nemusí být úplný svaz (například $(\mathbb{N}, \leq)$ není úplný svaz, neboť neexistuje supremum celé množiny $\mathbb{N}$).

Věta 1.5.1. Nechť $(G, \leq)$ je uspořádaná množina. Následující podmínky jsou ekvivalentní

→ $(G, \leq)$ je úplný svaz,

→ $(G, \leq)$ má nejmenší prvek a každá neprázdná podmnožina množiny G má v uspořádané množině $(G, \leq)$ supremum,

→ $(G, \leq)$ má největší prvek a každá neprázdná podmnožina množiny G má v uspořádané množině $(G, \leq)$ infimum.

Příklad 1.5.5. Svaz všech podgrup dané grupy G je dle předchozí věty úplný svaz, neboť má největší prvek (celou grupu G) a každá neprázdná množina podgrup má v tomto svazu infimum, kterým je průnik těchto podgrup. Rovněž svaz všech podsvazů (popřípadě svaz ideálů nebo svaz filtrů) daného svazu je úplný svaz. Díky analogickým větám o průnicích neprázdných množin určitých podstruktur lze totéž říci i o svazu všech podokruhů daného okruhu nebo o svazu jeho ideálu, o svazu všech podtěles daného tělesa nebo o svazu všech podprostorů daného vektorového prostoru.

Příklad 1.5.6. $(\mathbb{N} \cup \{\infty\}, \leq)$ je dle předchozí věty úplný svaz, neboť má největší prvek ∞ a každá neprázdná podmnožina množiny $\mathbb{N} \cup \{\infty\}$ má v $(\mathbb{N} \cup \{\infty\}, \leq)$ infimum.

Příklad 1.5.7. Ze svazu $(\mathbb{N}, |)$, který není úplný, lze doplněním nuly (která se stane jeho největším prvkem) utvořit úplný svaz $(\mathbb{N} \cup \{0\}, |)$.

Jak ukazuje následující věta, předchozí případy nebyly nijak výjimečné, vždy existuje způsob, jak doplnit svaz tak, aby se stal úplným.

Věta 1.5.2. Nechť G je svaz. Pak existuje úplný svaz I, který obsahuje podsvaz H, jenž je izomorfní se svazem G [2], [6].

1.5.1 Galoisova konexe

Definice: Jsou-li G a H uspořádané množiny a $f: G \to H$, $g: H \to G$ zobrazení, pak dvojice zobrazení (f, g) se nazývá *Galoisova konexe* mezi G a H, jestliže platí

→ zobrazení f a g jsou *antifonní*,

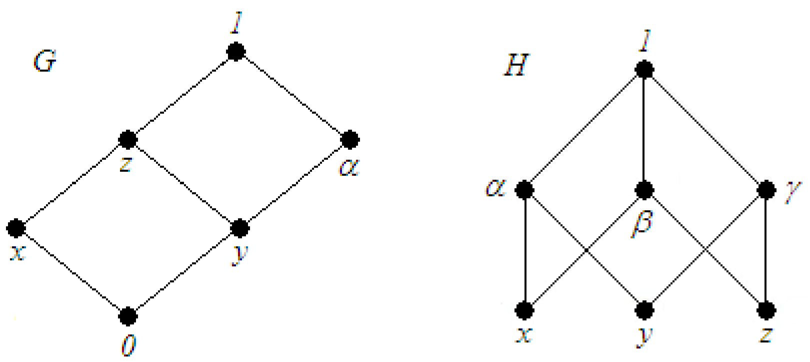
→ pro všechna $x \in G$ a $y \in H$ platí

$$x \leq g(f(x)),$$

$$y \leq f(g(y)).$$ [4]

Protože formální konceptuální analýza (dále jen FCA) je založena na Galoisových konexích mezi svazy všech podmnožin množiny objektů a atributů, uvedeme konkrétní příklad Galoisovy konexe mezi uspořádanými množinami.

Příklad 1.5.1.1. Uspořádané množiny G a H jsou reprezentovány Hasseovými diagramy viz následující obrázek.



Obr. 8. Hasseovy diagramy G a H

Galoisova konexe (f, g) mezi G a H je dána tabulkami zobrazení $f: G \to H$ a $g: H \to G$:

	0	x	y	z	α	*1*
f	*1*	α	*1*	α	γ	y

Tab. 1. Zobrazení $f: G \to H$

	x	y	z	α	β	γ	*1*
g	z	*1*	α	z	y	α	y

Tab. 2. Zobrazení $g: H \to G$

2 UZÁVĚROVÝ OPERÁTOR A VĚTA O PEVNÉM BODĚ

Některé důležité svazy splňují vedle identit obsažených v algebraické definici svazu a samozřejmě i všech formulí, které se z nich dají dokázat, ještě i některé další. Nejdůležitější jsou mezi těmito speciálními typy svazů svazy *modulární*, *distributivní* a *komplementární*. Proto se jimi v téhle kapitole budu podrobněji zabývat. Nejprve si tedy rozebereme modulární svazy, které dostáváme např. při studiu některých systémů podstruktur dané algebraické struktury a hrají důležitou roli i při studiu projektivních prostorů. Následně pak svazy distributivními, které jsou jejich zvláštním případem a svazy komplementární.

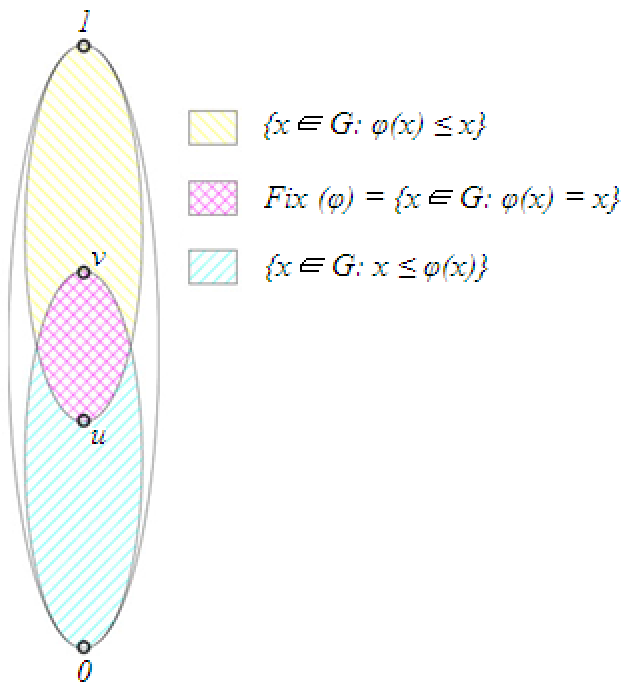
Věta 2.1. (Tarski) Nechť G je úplný svaz, $\varphi: G \to G$ izotonní zobrazení. Pak existuje prvek $x \in G$ tak, že $\varphi(x) = x$ (tj. x je pevný bod zobrazení φ).

Dále existuje největší pevný bod v zobrazení φ, pro který platí $v = \sup \{x \in G: x \leq \varphi x\}$ a nejmenší pevný bod u zobrazení φ, pro který platí $u=\inf \{x \in G: \varphi x \leq x\}$.

Tarského věta o pevném bodě má široké aplikace v *computer science*, zejména se používá při definicích sémantik programovacích jazyků. Dále platí, že množina pevných bodů každého izotonního zobrazení daného úplného svazu do sebe tvoří také úplný svaz. Další zajímavou vlastností Tarského věty o pevném bodě je skutečnost, že existenci pevného bodu každého izotonního zobrazení lze použít pro charakterizaci úplnosti svazů. Platí následující obrácení Tarského věty [1].

Jestliže každé izotonní zobrazení daného svazu do sebe má pevný bod, pak je daný svaz úplný. Odtud plyne zajímavá charakterizace úplnosti pro svazy.

Svaz G je úplný právě tehdy, když každé izotonní zobrazení svazu G do sebe má alespoň jeden pevný bod.



Obr. 9. Největší a nejmenší pevný bod v úplném svazu

Věta 2.2 (*Uzávěrový operátor*). Zobrazení $\varphi\colon G \to G$ uspořádané množiny G do sebe se nazývá uzávěrový operátor, jestliže pro každé $x, y \in G$ platí

$$x \leq \varphi(x),$$

$$x \leq y \text{ } implikuje \text{ } \varphi(x) \leq \varphi(y),$$

$$\varphi(x) = \varphi(\varphi(x)).$$ [6]

Věta 2.3 (*Charakterizace uzávěrových operátorů*). Libovolné zobrazení φ uspořádané množiny G do sebe je uzávěrový operátor právě tehdy, když pro všechna $x, y \in G$ platí ekvivalence

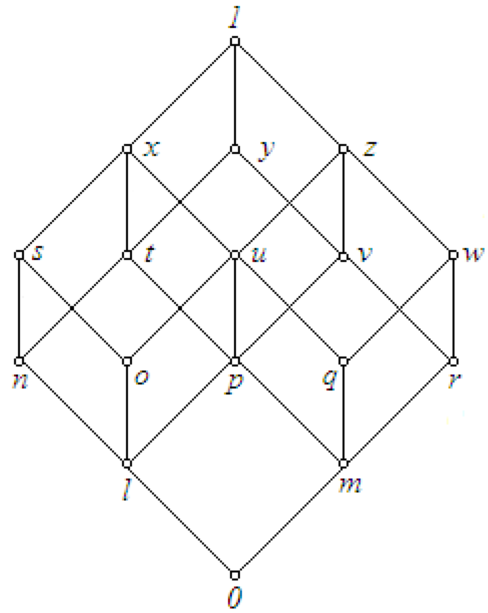
$$x \leq \varphi(y) \Leftrightarrow \varphi(x) \leq \varphi(y).$$

Teorie uzávěrových operátorů se využívá ve všech oblastech matematiky, zejména v topologii a je na ní založena definice konceptů ve formální konceptuální analýze. Pro úplné svazy platí následující věta [6].

Věta 2.4. Je-li φ uzávěrový operátor na úplném svazu G, pak množina všech pevných bodů $Fix\ (\varphi) = \{x \in G\colon \varphi(x) = x\}$ tvoří úplný svaz, ve kterém největší prvek je roven 1.

Na této větě je založena hlavní věta (*hlavní věta o konceptuálních svazech*) formální konceptuální analýzy, která tvrdí, že množina všech konceptů každého kontextu tvoří, vzhledem k uspořádání množinovou inkluzí úplný svaz, tzv. *konceptuální svaz*.

Příklad 2.1 (*Uzávěrový operátor*). Jako ukázku uzávěrových operátorů uvedeme konkrétní příklad na svazu $G = \{0, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 1\}$, jehož Hasseův diagram je znázorněn na obrázku níže.



Obr. 10. Hasseův diagram svazu G

Uzávěrové operátory f, g na svazu G jsou definovány předpisem v tabulce níže. Množiny pevných bodů těchto uzávěrových operátorů tvoří úplné svazy a platí

$$Fix\,(f) = \{p, t, u, v, 1\},$$

$$Fix\,(g) = \{p, t, u, v, x, y, z, 1\}.$$

Dále jsou v tabulce níže definovány složená zobrazení $f \circ g$ a $g \circ f$ z těchto dvou uzávěrových operátorů. Složení $f \circ g$ je uzávěrový operátor, protože splňuje všechny tři podmínky z **Věty 2.2.** a pro množinu pevných bodů tohoto složení platí

$$Fix\,(f \circ g) = \{p, t, u, v, 1\}.$$

Také složení $g \circ f$ tvoří uzávěrový operátor na G, protože jsou splněny všechny tři podmínky z **Věty 2.2.** Například pro prvek $l \in G$ platí

$$(g \circ f)(l) = p,$$

$$(g \circ f) \circ (f \circ g)(l) = g\Big(f\Big(g\big(f(e)\big)\Big)\Big) = g\big(f(p)\big) = p.$$

Množina pevných bodů tohoto složení

$$Fix\ (g \circ f) = \{p, t, u, v, 1\}$$

tvoří úplný svaz.

Složení dvou uzávěrových operátorů není obecně uzávěrový operátor, dokonce i v případě úplných svazů.

	0	***l***	***m***	***n***	***o***	***p***	***q***	***r***	***s***	***t***	***u***	***v***	***w***	***x***	***y***	***z***	***1***
$\boldsymbol{f}$	p	p	p	t	u	p	u	v	1	t	u	v	1	1	1	1	1
$\boldsymbol{g}$	p	p	p	t	u	p	u	v	x	t	u	v	z	x	y	z	1
$\boldsymbol{g \circ f}$	p	p	p	t	u	p	u	v	1	t	u	v	1	1	1	1	1
$\boldsymbol{f \circ g}$	p	p	p	t	u	p	u	v	x	t	u	v	z	1	1	1	1

Tab. 3. Definice uzávěrových operátorů f, g

2.1 Modulární svaz

Věta 2.1.1. Nechť $(G,\vee,\wedge)$ je libovolný svaz. Pak pro libovolné prvky $x, y, z \in G$ takové, že $x \geq z$, platí následující nerovnost

$$x \wedge (y \vee z) \geq (x \wedge y) \vee z.$$

Definice: Svaz $(G,\vee,\wedge)$ se nazývá *modulární*, platí-li

$$\forall x, y, z \in G; x \geq z \Rightarrow x \wedge (y \vee z) = (x \wedge y) \vee z.$$

I zde platí princip duality

$$\forall x, y, z \in G; x \leq z \Rightarrow x \vee (y \wedge z) = (x \vee y) \wedge z.$$

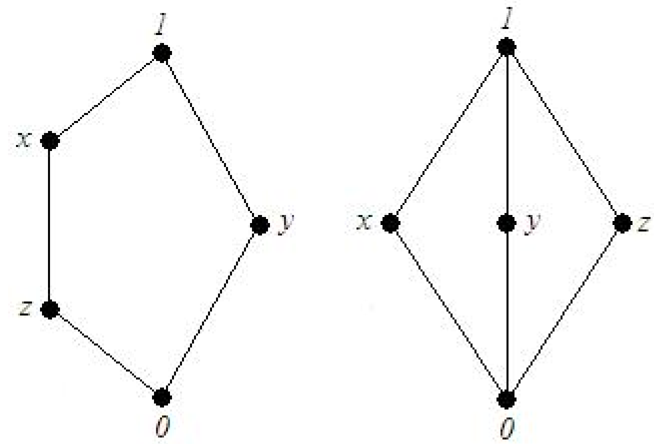
Příklad 2.1.1. Svaz $(2^X,\cup,\cap)$ všech podmnožin nějaké množiny X nebo libovolný řetězec.

Příklad 2.1.2. Svaz N_5, zvaný *pentagonální svaz*, není modulární, kdežto svaz M_5, zvaný *diamantový svaz*, modulární je (viz obrázek níže). Označme $0 < z < x < 1$, prvky, které jsou v Hasseově diagramu svazu N_5 nakresleny nad sebou vlevo, a y jeho prvek, pak nerovnost

$$(x \wedge y) \vee z = 0 \vee z = z < x = x \wedge 1 = x \wedge (y \vee z)$$

ukazuje, že svaz N_5 není modulární.

Nyní probírkou všech možností dokažme, že svaz M_5 je modulární. Označme 0 nejmenší a 1 největší prvek tohoto svazu. Nechť tedy $x, y, z \in M_5$ jsou libovolné takové, že $z \leq x$. Jestliže $x = z$, plyne modulární rovnost z absorpčních zákonů. Jestliže $z < x$, pak na Hasseově diagramu svazu M_5 vidíme, že buď $z = 0$ nebo $x = 1$. V obou případech je modulární rovnost zřejmá [6].



Obr. 11. Svaz N_5 (Pentagonální svaz) a svaz M_5 (Diamantový svaz)

Věta 2.1.2. Každý podsvaz modulárního svazu je modulární.

Příklad 2.1.3. Svaz všech podprostorů daného vektorového prostoru V nad tělesem T je podle předchozí věty modulární. Je totiž podsvazem modulárního svazu všech podgrup grupy vektoru V, k tomu si stačí uvědomit, že každý podprostor je podgrupou, a ověřit, že infima i suprema se ve svazu všech podprostorů počítají stejně jako ve svazu podgrup. Infimem je množinový průnik a supremem součet podprostorů.

Věta 2.1.3. Každý distributivní svaz je modulární.

Platí tedy, že třída distributivních svazů je podtřídou třídy modulárních svazů. Přitom modulární svaz nemusí být distributivní, proto se uvedené třídy svazů nerovnají.

Věta 2.1.4. Jestliže $\mathcal{G} = (G, \cdot)$ je grupa a $S_N(\mathcal{G})$ je množina jejích normálních podgrup, pak svaz $(S_N(\mathcal{G}), \subseteq)$ je modulární.

Věta 2.1.5. Svaz $(G, \vee, \wedge)$ je modulární právě tehdy, když platí

$$\forall x, y, z \in G; x \wedge \big(y \vee (x \wedge z)\big) = (x \wedge y) \vee (x \wedge z).$$

Součin modulárních svazů je modulární svaz. Homomorfní obraz modulárního svazu je modulární svaz.

Věta 2.1.6. Jestliže $(G,\vee,\wedge)$ je modulární svaz, $x, y, z \in G$ a $x \geq y$, pak

$$x \vee z = y \vee z, x \wedge z = y \wedge zc \Rightarrow x = y.$$

Věta 2.1.7. Svaz G je modulární právě tehdy, když neobsahuje žádný podsvaz izomorfní se svazem N_5.

Každý podsvaz modulárního svazu je modulární, zatímco N_5 modulární není [1], [4], [6].

2.2 Semimodulární svaz

Věta 2.2.1. Svaz G je *semimodulární*, jestliže pro každé $x, y \in G$ platí

$$x \wedge y \leq x \Rightarrow y \leq x \vee y.$$

Věta 2.2.2. Svaz G je *spodně modulární*, jestliže pro každé $x, y \in G$ platí

$$y \leq x \vee y \Rightarrow x \wedge y \leq x.$$

Konečný svaz je modulární tehdy a jen tehdy, pokud je semimodulární a zároveň spodně semimodulární. Tedy platí-li obě výše zmíněné implikace současně

$$x \wedge y \leq y \Leftrightarrow y \leq x \vee y \text{ [5]}.$$

2.3 Distributivní svaz

Věta 2.3.1. Nechť $(G,\vee,\wedge)$ je libovolný svaz. Pak pro libovolné prvky $x, y, z \in G$ platí následující nerovnosti

$$x \vee (y \wedge z) \leq (x \vee y) \wedge (x \vee z),$$

$$x \wedge (y \vee z) \geq (x \wedge y) \vee (x \wedge z).$$

Druhá nerovnost je duální k první, tedy podle principu duality teorie svazů také platí.

Definice: Svaz $(G,\vee,\wedge)$ se nazývá *distributivní*, platí-li

$$\forall x, y, z \in G; x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z),$$

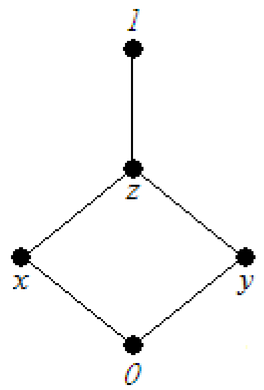
$$\forall x, y, z \in G; x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

Jestliže je ve svazu G splněna první podmínka, pak je v něm splněna také podmínka druhá a obráceně. Duální svaz k distributivnímu svazu je opět distributivní.

Příklad 2.3.1. Svaz všech podmnožin nějaké množiny nebo libovolný řetězec.

Příklad 2.3.2. Nechť M je libovolná množina. Pak ve svazu $(G(M), \subseteq)$ platí, že průseky se rovnají průnikům a spojení se rovnají sjednocením, proto ze známých vlastností množinových operací dostáváme, že svaz $G(M)$ je distributivní.

Příklad 2.3.3. Svaz s Hasseovým diagramem na následujícím obrázku je distributivní.



Obr. 12. Hasseův diagram distributivního svazu

Věta 2.3.2. Jestliže G je distributivní svaz, pak také každý jeho podsvaz je distributivním svazem.

Věta 2.3.3. Jestliže ve svazu G platí pro libovolné prvky $x, y, z \in G$ implikace

$$x \wedge y = x \wedge z, x \vee y = x \vee z \Rightarrow y = z,$$

pak G je distributivní svaz.

Věta 2.3.4. Každý distributivní svaz je modulární.

Věta 2.3.5. Součin distributivních svazů je distributivní svaz. Homomorfní obraz distributivního svazu je distributivní svaz.

Věta 2.3.6. Svaz G je distributivní právě tehdy, když žádný podsvaz v G není izomorfní se svazem M_5 (diamantový svaz) ani se svazem N_5 (pentagonální svaz).

Na závěr zde ještě uvedu charakterizaci konečných distributivních svazů.

Definice: Prvek x svazu G se nazývá $\vee$ - nedosažitelný, jestliže pro každé $y, z \in G$ takové, že $x = y \vee z$, platí $x = y$ nebo $x = z$.

Prvek x svazu G je tedy $\vee$ - nedosažitelný, jestliže není supremem žádných dvou prvků ostře menších než on, tj. neexistují $y, z \in G$ splňující $y < x, z < x, x = y \vee z$. Ekvivalentně lze tuto podmínku vyjádřit také takto: prvek x svazu G je $\vee$ - nedosažitelný, jestliže pro každé $y, z \in G$ takové, že $y < z$ a současně $z < x$, platí $y \vee z < x$. Odtud se snadno dokáže indukcí, že takový prvek není supremem ani žádné neprázdné konečné množiny prvku ostře menších než on.

Množinu všech $\vee$ - nedosažitelných prvků svazu G označíme $J(G)$.

Věta 2.3.7. V konečném distributivním svazu G je libovolný prvek x roven supremu množiny všech $\vee$ - nedosažitelných prvků, které neostře převyšuje, tj.

$$x = \bigvee\nolimits_{y \in J(G), y \leq x}, b = \bigvee(x \downarrow\cap J(G)).$$

Nechť $(G, \leq)$ je uspořádaná množina. Množina $(H \subseteq G)$ se nazývá (dolů) dědičná, pokud pro každý prvek $(y \in H)$ a každý $(x \in G)$, $x \leq y$, platí $(x \in H)$.

Množina $(H \subseteq G)$ je tedy dědičná, jestliže s každým svým prvkem obsahuje všechny prvky množiny G, které jsou ještě menší. Pomocí této vlastnosti můžeme charakterizovat ideály svazu: jsou to právě dědičné podsvazy. Připomeňme, že na svazy se můžeme dívat jako na uspořádané množiny a že dva svazy jsou izomorfní, právě když jsou izomorfní jako uspořádané množiny.

Množinu všech neprázdných dědičných podmnožin uspořádané množiny G značíme $D(G)$.

Věta 2.3.8. Pro konečný distributivní svaz G uvažme množinu $J(G)$ všech V - nedosažitelných prvků svazu G spolu s uspořádáním, které na $J(G)$ indukuje uspořádání svazu G. Pak uspořádaná množina $(D(J(G)), \subseteq)$ je izomorfní se svazem G (chápaným jako uspořádaná množina).

Věta mimo jiné říká, že je-li G konečný distributivní svaz, pak i $D(J(G))$ je konečný distributivní svaz. Protože sjednocení i průnik dědičných množin je opět dědičná množina, jsou operacemi suprema a infima v $D(J(G))$ právě množinový průnik a sjednocení. Je tedy $D(J(G))$ podsvazem svazu všech podmnožin množiny $J(G)$.

Každý konečný distributivní svaz je izomorfní s některým podsvazem svazu všech podmnožin nějaké konečné množiny.

Podle předchozího důsledku každý konečný distributivní svaz můžeme chápat jako inkluzí uspořádaný systém množin, který je uzavřený na průniky a sjednocení. Protože naopak každý inkluzí uspořádaný systém množin, který je uzavřený na průniky a sjednocení, je zřejmě distributivním svazem, dostali jsme tak slíbenou charakterizaci konečných distributivních svazů [1], [4], [6].

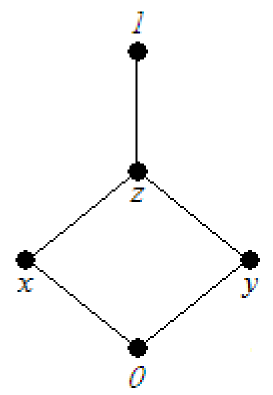
2.4 Komplementární svaz

Nechť $(G, \vee, \wedge)$ je svaz s nejmenším prvkem 0 a největším prvkem 1. Nechť $x \in G$. Pak prvek $y \in G$ se nazývá komplement prvku x, platí-li

$$x \vee y = 1,$$

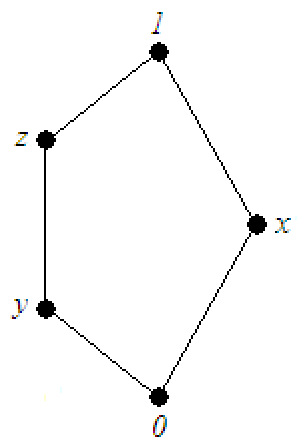
$$x \wedge y = 0.$$

Příklad 2.4.1. Nechť G je svaz s diagramem na obrázku níže. Pak 0 je komplementem 1 a 1 je komplementem 0, ale žádný další prvek z G už komplement v G nemá.



Obr. 13. Diagram svazu

Příklad 2.4.2. Ve svazu N_5 má každý prvek aspoň jeden komplement. Přitom prvek x má dva různé komplementy y, z.



Obr. 14. Diagram svazu

Věta 2.4.1. Jestliže svaz G je distributivní, pak každý jeho prvek má nejvýše jeden komplement.

Definice: Svaz G se nazývá komplementární, jestliže má nejmenší a největší prvek (tj. G je ohraničený) a jestliže každý prvek $x \in G$ má v G aspoň jeden komplement.

Booleovým svazem pak rozumíme každý svaz $(H,\vee,\wedge)$, který je distributivní a komplementární.

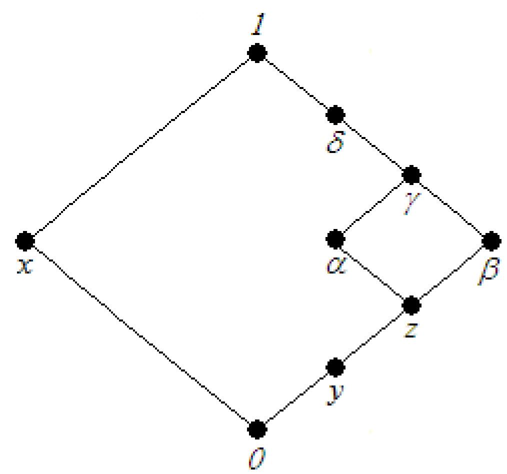
Svaz se nazývá jednoznačně komplementární, jestliže každý jeho prvek má právě jeden komplement. Proto je každý Booleův svaz je jednoznačně komplementární.

Definice: Svaz $(G,\vee,\wedge)$ se nazývá relativně komplementární, jestliže pro libovolné prvky $x, y, z \in G$ takové, že $x \leq y \leq z$, existuje prvek $\alpha \in G$ takový, že

$$z \wedge \alpha = x,$$

$$z \vee \alpha = y.$$

Komplementární svaz ale nemusí být relativně komplementární. Příkladem takového svazu je svaz s následujícím Hasseovým diagramem [4].



Obr. 15. Diagram komplementárního svazu

II. PRAKTICKÁ ČÁST

3 FORMÁLNÍ KONCEPTUÁLNÍ ANALÝZA

Formální konceptuální analýza (dále jen FCA) neboli metoda konceptuálních svazů má mnoho využití v různých oblastech. Jedná se o metodu datové analýzy (analýzy tabulkových dat), která vychází z teorie svazů zabývající se uspořádanými množinami, kde ke každým dvěma prvkům existuje supremum i infimum. Tato část je rozebrána v kapitole 1.

Pokud lidé definují své znalosti a vědomosti o okolním světě, popisují jej nejčastěji pomocí objektů a jejich atributů. Obecně se tedy snaží formulovat různě složitá tvrzení o tom, že některé objekty mají některé atributy. Mezi těmito objekty a atributy přitom existuje základní vztah vyjádřený pomocí slovesa mít. Pro daný objekt a daný atribut totiž platí, že objekt má daný atribut nebo jej nemá, popřípadě má atribut pouze do určité míry, s jistou hodnotou atd. Tento vztah nejlépe znázorňuje tabulka (matice), ve které řádky představují objekty a sloupce atributy, přičemž konkrétní položka tabulky odpovídající objektu x a atributu y podává informaci o tom, zda a popřípadě s jakou hodnotou má objekt x atribut y.

	y_1	.	.	y_j	.	.	y_l
x_1				.			
.				.			
.				.			
x_i	.	.		$G(x_i, y_j)$		.	.
.				.			
.				.			
x_k				.			

Tab. 4. Tabulková data s objekty x_i a atributy y_j.

Tabulková data jsou elementárním a neodmyslitelným zdrojem informací pro nejrůznější metody analýzy a zpracování dat.

Jednou z metod analýzy tabulkových dat je formální konceptuální analýza neboli metoda konceptuálních svazů. Tato metoda má široké spektrum využití v nejrůznějších oblastech. Vychází z teorie svazů zabývající se uspořádanými množinami, přičemž uživateli nabízí netriviální informace o vstupních datech, které mohou být využitelné přímo (upozorňuje na nové poznatky o vstupních datech, které nemusí být při prvním pohledu na vstupní data patrné), popřípadě mohou být využitelné při dalším zpracování dat.

Formální konceptuální analýza je přitom částí aplikované matematiky, která dokáže definovat a zachytit objekty a jejich vlastnosti neboli běžně se objevující údaje v mnoha oblastech lidské činnosti. Slouží jako prostředek k uspořádání pojmů a zachycení grafického výstupu tak, aby z něj bylo patrné, které objekty jsou obecnější než jiné, a hlavně jak spolu tyto objekty vzájemně souvisí.

Pomocí formální konceptuální analýzy je možné elementárně získat tzv. *konceptuální svaz* nebo tzv. *atributové implikace*. Rozdíl mezi těmito dvěma výstupy spočívá v tom, že konceptuální svaz představuje hierarchicky uspořádanou množinu určitých shluků dosazených ze vstupní tabulky dat, tzv. *formálních konceptů*, zatímco na druhé straně atributové implikace zobrazují vzájemné závislosti mezi jednotlivými vlastnostmi v tabulce dat. Nespornou výhodou konceptuálního svazu, na rozdíl od jiných analytických metod, je možnost nahlížet na nalezené koncepty jako na celek, neboť konceptuální svaz se tvoří nad celým vstupním kontextem a data stále obsahují všechny podrobnosti týkající se zadaných souvislostí.

Při práci s formální konceptuální analýzou se pro zjednodušení předpokládá, že atributy ve vstupních datech jsou bivalentní, tzn., že mohou nabývat pouze hodnot ano nebo ne. Pro každý atribut y a každý uvažovaný objekt x tedy platí, že x má y nebo x nemá y. Z tohoto vyplývá, že jedinými možnými hodnotami jsou buď 0, nebo 1, přičemž hodnota 0 znamená nepravdu a hodnota 1 značí pravdu. Objekty mohou ale nabývat kromě bivalentních logických atributů i dalších hodnot atributů, kdy objekt x má například vlastnost y s hodnotou w. V tomto případě se jedná o složitější situaci, kterou jé možné zjednodušit tzv. *konceptuálním škálováním*.

G	y_1	y_2	y_3	y_4
x_1	1	1	0	0
x_2	0	1	1	0
x_3	0	0	1	1

Tab. 5. Bivalentní logické atributy

Formální konceptuální analýza přesně definuje výraz „pojem". Z lidského hlediska lze pod tímto termínem chápat skupinu objektů, které mají jisté společné vlastnosti. Matematicky to pak lze vyjádřit jako uspořádanou dvojici (x, y), kde x představuje skupinu objektů a y je množina atributů, které objektům patří. Taková uspořádaná dvojice je pojmem tehdy, když x je množina všech objektů, které sdílejí všechny atributy z y, a současně y je množina všech atributů, které jsou společné všem objektům z x. Formální konceptuální analýza takto definovaný pojem nazývá *konceptem*, popř. *formálním konceptem,* a v tabulkových datech odpovídá maximálním obdélníkům vyplněným jedničkami. Koncepty se navzájem liší svou obecností, jeden koncept je obecnější než koncept jiný. Řekněme tedy, že koncept (x_1, y_1) je podpojmem konceptu (x_2, y_2), (tj. první koncept je nejvýše tak obecný jako druhý), pokud platí, že každý objekt z x_1 patří do x_2 nebo, což je ekvivalentní, že každý atribut z y_2 patří do y_1. Tuto podmínku obecně značíme $(x_1, y_1) \leq (x_2, y_2)$. Například kočka je podpojmem konceptu savec. Tenhle definovaný vztah tedy umožňuje množinu všech konceptů uspořádat podle jejich obecnosti. Množina všech konceptů uspořádaných podle jejich obecnosti se pak nazývá *konceptuální svaz*.

Jednotlivé atributy v tabulce vstupních dat jsou mezi sebou závislé a právě tyto závislosti jsou popsány v tzv. *atributových implikacích* ve tvaru atributy $y_1, \dots, z_1$ implikují atributy $y_2, \dots, z_2$. Formální zápis pak vypadá následovně $\{y_1, \dots, z_1\}$ $\{y_2, \dots, z_2\}$. To v praxi znamená, že každý koncept s vlastnostmi $y_1, \dots, z_1$, má tím pádem i vlastnosti $y_2, \dots, z_2$. Ve FCA většinou hledáme nějakou množinu implikací, které nejsou nadbytečné, tzn., že *nejsou triviální* a na první pohled zřejmé. Z těchto implikací lze všechny ostatní logicky odvodit. Podrobněji se danými pojmy budu zabývat v následujících kapitolách.

3.1 Historie FCA

Základy formální konceptuální analýzy položil již v roce 1982 Rudolf Wille, který při svých výzkumech navázal na práce Garretta Birkhoffa o teorii svazu a uspořádání. Rudolf Wille se zabýval konceptuálními svazy v rámci svého programu tzv. *restrukturalizace teorie svazu* a snažil se přiblížit teorii svazu k praktickému využití. Ve své knize popsal teoretické znalosti a veškerá fakta týkající se formální konceptuální analýzy.

Příspěvky k FCA a konceptuálním svazům je možné rovněž najít v různých matematických, informatických a inženýrských časopisech, jsou také přednášeny na různých konferencích věnujících se analýze dat. Za zmínku stojí zejména konference General Aplgebra and Its Applications a nověji konference Conceptual Structures.

3.2 Vymezení základních pojmů a definice FCA

Tato část diplomové práce definuje základní termíny z odborné literatury a popisuje základní pojmy, které řeší problematiku formální konceptuální analýzy.

3.2.1 Formální kontext, indukované Galoisovy konexe

Definice: Formální kontext je trojice $< X, Y, G >$, kde X a Y jsou neprázdné množiny a G je binární relace mezi X a Y, tedy $G \subseteq X \times Y$. Prvky x z množiny X se nazývají *objekty*, zatímco prvky y z množiny Y jsou tzv. *atributy*. Zápis $< x, y > \in G$ pak vyjadřuje, že objekt x má atribut y. Formální kontext tedy reprezentuje výše zmíněná tabulková objekt – atributová data.

Každý kontext $< X, Y, G >$ indukuje zobrazení předpisem $\uparrow : 2^x \to 2^y\ a\ \downarrow : 2^y \to 2^x$, kde:

$$A^{\uparrow} = \{y \in Y | \forall x \in A: \langle x, y \rangle \in G\} \text{ pro } A \subseteq X$$

a

$$B^{\downarrow} = \{x \in X | \forall y \in B: \langle x, y \rangle \in G\} \text{ pro } B \subseteq Y.$$

Pomocí symbolu $\uparrow$ lze přiřadit podmnožiny Y do podmnožiny X. $A^{\uparrow}$ je pak množina atributů sdílených všemi objekty z A. Stejný princip platí i obráceně. Operátor $\downarrow$ přiřadí podmnožiny X do podmnožiny Y. $B^{\uparrow}$ je pak množina objektů sdílených všemi atributy z B.

Pokud není kontext příliš rozsáhlý, lze ho přehledně znázornit pomocí tabulky, která se označuje jako tabulka kontextová.

Příklad 3.2.1.1.

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

Tab. 6. Příklad kontextové tabulky

- $\{x_2\}^{\uparrow} = \{y_2, y_3, y_4\}, \{x_2, x_3\}^{\uparrow} = \{y_3, y_4\}$,
- $\{x_2, x_3, x_5\}^{\uparrow} = \emptyset$,
- $X^{\uparrow} = \emptyset, \emptyset^{\uparrow} = Y$,
- $\{y_1\}^{\downarrow} = \{x_1, x_3, x_4, x_5\}, \{y_1, y_2\}^{\downarrow} = \{x_1, x_4\}$,
- $\{y_2, y_3\}^{\downarrow} = \{x_1, x_2, x_4\}, \{y_2, y_3, y_4\}^{\downarrow} = \{x_1, x_2, x_4\}$,
- $\emptyset^{\downarrow} = X, Y^{\downarrow} = \{x_1\}$.

Definice: Zobrazení $f: 2^X \to 2^Y$ a $g: 2^Y \to 2^X$ tvoří tzv. *Galoisovu konexi* mezi množinami X a Y, pokud pro A, A_1, $A_2 \subseteq X$ a $B, B_1, B_2 \subseteq Y$ platí $A_1 \subseteq A_2$ implikuje $f(A_2) \subseteq f(A_1)$; $B_1 \subseteq B_2$ implikuje $g(B_2) \subseteq g(B_1)$; $A \subseteq g\big(f(A)\big)$; $B \subseteq f\big(g(B)\big)$ [8].

Věta 3.2.1.1. Pro binární relaci $G \subseteq X \times Y$ tvoří indukovaná zobrazení ↑ G a ↓ G Galoisovu konexi mezi X a Y. Naopak, tvoří-li f a g Galoisovu konexi mezi X a Y, existuje binární relace $G \subseteq X \times Y$ tak, že $f =$ ↑ G a $g =$ ↓ G. Tím je dán vzájemně jednoznačný vztah mezi Galoisovými konexemi mezi X a Y a binárními relacemi mezi X a Y [8].

3.2.2 Formální koncept

Definice: Formální koncept v kontextu $\langle X, Y, G\rangle$ je dvojice $\langle A, B\rangle$, kde $A \subseteq X$ a $B \subseteq Y$ jsou takové, že $A^{\uparrow} = B$ a $B^{\downarrow} = A$.

A je přitom množina objektů nazývaná *extent* a B je množina atributů označovaná *intent* daného formálního konceptu $\langle A, B\rangle$ v kontextu $\langle X, Y, G\rangle$, kde A obsahuje právě všechny objekty sdílející atributy z B a B obsahuje pouze atributy, které jsou sdíleny všemi objekty z A.

Jak již bylo jednou v textu zmíněno, uspořádaná dvojice (A, B) je formální koncept pouze tehdy, když A je množina všech objektů, které sdílejí všechny atributy z B, a zároveň B je množina všech atributů, které jsou společné všem objektům z A.

Matematicky je koncept pevným bodem Galoisovy konexe dané $\uparrow$ a $\downarrow$.

Množinu všech formálních konceptů v $\langle X, Y, G\rangle$ značíme $\mathcal{B}(X, Y, G)$, tj.

$$\mathcal{B}(X, Y, G) = \left\{(A, B) | A \subseteq X, B \subseteq Y, A^{\uparrow} = B, B^{\downarrow} = A\right\} \text{ [8].}$$

Příklad 3.2.2.1. Formální koncept pro předchozí kontextovou tabulku je zvýrazněn níže:

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

Tab. 7. Příklad formálního konceptu

Zvýrazněný obdélník reprezentuje následující formální kontext:

$$\langle A_1, B_1\rangle = \langle\{x_1, x_2, x_3, x_4\}, \{y_3, y_4\}\rangle$$

protože

$$\{x_1, x_2, x_3, x_4\}^{\uparrow} = \{y_3, y_4\} \text{ a } \{y_3, y_4\}^{\downarrow} = \{x_1, x_2, x_3, x_4\}.$$

Tabulka ovšem obsahuje i jiné formální koncepty. Tři z nich jsou znázorněny níže:

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

Tab. 8. Doplnění zbylých konceptů

Jsou jimi:

- $\langle A_2, B_2\rangle = \langle\{x_1, x_2, x_4\}, \{y_2, y_3, y_4\}\rangle$
- $\langle A_3, B_3\rangle = \langle\{x_1, x_4\}, \{y_1, y_2, y_3, y_4\}\rangle$

$$\langle A_4, B_4\rangle = \langle\{x_3, x_{4,}, x_5\}, \{y_1\}\rangle.$$

3.2.2.1 *Formální koncept jako maximální obdélník*

Formální koncept lze také jednoduše vyjádřit jako největší obdélník v kontextové tabulce.

Definice: Obdélník v $\langle X, Y, G\rangle$ je pár $\langle A, B\rangle$ takový, že $A \times B \subseteq G$, tj. pro každé $x \in A$ a $y \in B$ máme $\langle X, Y\rangle \in$ G. Pro obdélníky $\langle A_1, B_1\rangle$ a $\langle A_2, B_2\rangle$ je dáno $\langle A_1, B_1\rangle \sqsubseteq \langle A_2, B_2\rangle$ jestliže $A_1 \subseteq A_2$ a $B_1 \subseteq B_2$.

G	y_1	y_2	y_3	y_4
x_1	1	1	1	1
x_2	0	1	1	1
x_3	1	0	1	1
x_4	1	1	1	1
x_5	1	0	0	0

Tab. 9. Formální koncept (největší obdélník) dané kontextové tabulky

V této tabulce není obdélník $\langle\{x_1, x_2, x_3\}, \{y_3, y_4\}\rangle$ největší možný, naopak obdélník $\langle\{x_1, x_2, x_3, x_4\}, \{y_3, y_4\}\rangle$ je již maximální obdélník dané kontextové tabulky.

$\langle A, B\rangle$ je formální koncept $\langle X, Y, G\rangle$ jestliže je $\langle A, B\rangle$ maximální obdélník v $\langle X, Y, G\rangle$ [8].

3.2.3 Konceptuální svaz

Konceptuální svaz představuje seskupení všech formálních konceptů existujícího formálního kontextu.

Definice: Konceptuální svaz je množina $\mathcal{B}(X,Y,G)$ spolu s relací $\leq$ definovanou na $\mathcal{B}(X,Y,G)$ předpisem

$(A_1,B_1) \leq (A_2,B_2)$ právě když $A_1 \subseteq A_2$ (nebo, ekvivalentně, $B_2 \subseteq B_1$).

Abychom s danými informacemi mohli dále pracovat, označíme $Int(I) = \{B \subseteq Y | \langle A,B\rangle \in \mathcal{B}(X,Y,G)\}$ pro nějakou $A \subseteq X$ jako množinu obsahů všech konceptů z $\mathcal{B}(X,Y,G)$. Přitom platí, že $B \subseteq Y$ je obsahem nějakého konceptu z $\mathcal{B}(X,Y,G)$. Stejný postup bude platit i v případě $Ext(G)$, který značí rozsahy konceptů z $\mathcal{B}(X,Y,G)$.

Vzájemný vztah $\leq$ tedy určuje podpojem – nadpojem. S tím souvisí i hlavní věta o konceptuálních svazech, která definuje strukturu $\mathcal{B}(X,Y,G)$ [8].

Věta 3.2.3.1. Hlavní věta o konceptuálních svazech: Mějme formální kontext $\langle X,Y,G\rangle$. $\mathcal{B}(X,Y,G)$ je vzhledem k $\leq$ úplný svaz, ve kterém jsou infima a suprema definována předpisy:

$$\bigwedge_{j\in J}\langle A_j,B_j\rangle = \langle\bigcap_{j\in J} A_j, \left(\bigcap_{j\in J} A_j\right)^{\uparrow}\rangle = \langle\bigcap_{j\in J} A_j, \left(\bigcup_{j\in J} B_j\right)^{\downarrow\uparrow}\rangle,$$

$$\bigvee_{j\in J}\langle A_j,B_j\rangle = \langle\left(\bigcap_{j\in J} B_j\right)^{\downarrow}, \bigcap_{j\in J} B_j\rangle = \langle\left(\bigcup_{j\in J} A_j\right)^{\uparrow\downarrow}, \bigcap_{j\in J} B_j\rangle.$$

Daný úplný svaz $\mathcal{V} = \langle V, \sqsubseteq\rangle$ je izomorfní s $\mathcal{B}(X,Y,G)$, právě když existují zobrazení $\gamma: X \to V$, $\mu: Y \to V$, pro která je $\gamma(X)$ supremálně hustá ve $\mathcal{V}$, $\mu(Y)$ infimálně hustá ve $\mathcal{V}$ a $\langle x,y\rangle \in G$ platí právě když $\gamma(x) \leq \mu(y)$ (pro každé $x \in X, y \in Y$).

Množina $K \subseteq \mathcal{V}$ je tedy supremálně hustá ve $\mathcal{V}$, právě když pro každý $v \in \mathcal{V}$ existuje $K_v \subseteq K$ tak, že v je supremem množiny K_v. Podobně pro infimálně hustou podmnožinu [8].

3.2.4 Atributové implikace

Atributové implikace poskytují především primární informace o závislostech v kontextových tabulkách.

Atributová implikace nad množinou Y atributů je výraz tvaru $A \Rightarrow B$, kde $A, B \subseteq Y$.

Definice: Pro implikaci $A \Rightarrow B$ a množinu $C \subseteq Y$ říkáme, že $A \Rightarrow B$ platí v C, popř. že C je modelem $A \Rightarrow B$, jestliže platí, že pokud $A \subseteq C$, pak i $B \subseteq C$. Obecněji, pro množinu $\mathcal{M} \subseteq 2^Y$ množin atributů a množinu $T = \{A_j \Rightarrow B_j;\ j \in J\}$ implikací říkáme, že T platí v M, popř. že M je modelem T, jestliže $A_j \Rightarrow B_j$ platí v C pro každé $C \in \mathcal{M}$ a $A_j \Rightarrow B_j \in T$.

Věta 3.2.4.1. Atributová implikace platí v $\langle X, Y, G\rangle$, právě když platí v $\mathcal{B}(X, Y, G)$.

Definice: Implikace $A \Rightarrow B$ (sémanticky) plyne z množiny T implikací (zapisujeme $T \vDash A \Rightarrow B$), jestliže $A \Rightarrow B$ platí v každé $C \subseteq Y$, ve které platí T. Množina T implikací se nazývá

- uzavřená, jestliže obsahuje každou implikaci, která z ní plyne;
- neredundantní, jestliže žádná implikace z T neplyne z ostatních (tj. nikdy není $T - \{A \Rightarrow B\} \vDash A \Rightarrow B$).

Množinu T implikací kontextu $\langle X, Y, G\rangle$ označujeme jako úplnou, jestliže z ní plyne každá implikace kontextu $\langle X, Y, G\rangle$. Báze je úplná a neredundantní množina implikací daného kontextu.

Ve formální konceptuální analýze obvykle hledáme nějakou množinu implikací, které nejsou nadbytečné, tzn., nejsou triviální a na první pohled zřejmé. Z těchto implikací lze všechny ostatní logicky odvodit. Nejčastěji vynecháváme implikace typu $A \Rightarrow B$ takové, že $B \subseteq A$, nebo implikace, které z ostatních plynou zcela přirozeně. Je nutné neustále kontrolovat, zda aktuální množina je stále úplná (všechny implikace z kontextu z ní plynou).

Věta 3.2.4.2. Množina T implikací je uzavřená, právě když pro každé $A, B, C, D \subseteq Y$ platí:

- $A \Rightarrow A \in T$;
- pokud $A \Rightarrow B \in T$, pak $A \cup C \Rightarrow B \in T$;

- pokud $A \Rightarrow B \in T$ a $B \cup C \Rightarrow D \in T$, pak $A \cup C \Rightarrow D \in T$ [8].

3.2.5 Vícehodnotové kontexty a konceptuální škálování

Některé konkrétní situace neumožňují vyjádřit vstupní data pouze pomocí bivalentních logických hodnot, ale k jejich popisu je nutné využít i jiné atributy, což vedlo k rozšíření formálních kontextů a k vytvoření kontextů vícehodnotových (many valut contexts).

Definice: Vícehodnotový kontext je čtveřice $\langle X, Y, Z, G \rangle$, kde $G \subseteq X \times Y \times Z$ je ternární relace taková, že pokud $\langle x, y, z \rangle \in G$ a $\langle x, y, w \rangle \in G$, pak $z = w$.

Prvky množin X, Y se stejně jako u formálních kontextů nazývají objekty a atributy (zde jsou na rozdíl od formálních kontextů vícehodnotové). Prvky množiny Z pak představují hodnoty těchto atributů. Skutečnost $\langle x, y, z \rangle \in G$ představuje situaci, že objekt x má atribut y s hodnotou z, což můžeme napsat i takto $y(x) = z$.

Formální konceptuální analýza přistupuje k rozboru vícehodnotových kontextů následujícím způsobem. Vícehodnotový kontext je potřeba nejprve pomocí vhodného tzv. *konceptuálního škálování* převést na základní kontext, který je teprve možné důkladně zkoumat.

Definice: Škála (scale) pro atribut y vícehodnotového kontextu je kontext $S_y = \langle X_y, Y_y, G_y \rangle$, pro který $y(X) \subseteq X_y$ (kde $y(X) = \{y(x) | x \in X\}$). Prvky množin X_y a Y_y se nazývají škálové hodnoty a škálové atributy.

Škálou pro daný atribut vícehodnotového kontextu může být naprosto libovolný kontext, ovšem podmínkou je, že musí splňovat podmínky předchozí definice. Může se také stát, že některý kontext je vhodnější než kontext jiný, protože lépe vyjadřuje smysl daného atributu. Pro takové atributy, které se ve vícehodnotových kontextech poměrně běžně vyskytují, existuje celá řada standardních škál, mezi které patří například škála nominální, ordinální, interordinální, biordinální, dichotomická, atd.

Převádění vícehodnotového kontextu na základní kontext se provádí pomocí tzv. *jednoduchého škálování (plainscaling)*.

Definice: Je-li $\langle X, Y, Z, G \rangle$ vícehodnotový kontext a jsou-li $S_y(y \in Y)$ škály, pak kontext odvozený jednoduchým škálováním je kontext $\langle X, K, J \rangle$, kde

- $N = \cup_{y \in Y} \dot{Y}_y$ ($\dot{Y}_y = \{y\} \times Y_y$);
- $\langle x, \langle y, k \rangle \rangle \in J$ právě když $y(x) = z$ a $\langle z, k \rangle \in I_y$.

Z definice jasně vyplývá, že objekty vícehodnotového a základního kontextu jsou naprosto shodné, zatímco atributy základního kontextu získáme pomocí disjunktního sjednocení atributů jednotlivých škál. V praxi se jednoduché škálování provádí tak, že v tabulce místo sloupce s označením y vložíme $|Y_y|$ sloupců označených atributy z Y_y a každou hodnotu $y(x)$ z vícehodnotového kontextu nahradíme řádkem škály S_y příslušným objektu x [8].

Příklad 3.2.5.1. V tabulce níže je zobrazen příklad vícehodnotových kontextů.

	y_1	y_2	y_3	y_4
x_1	12	1	1	0
x_2	3	1	8	1
x_3	7	0	27	1
x_4	24	1	15	0

Tab. 10. Vícehodnotové kontexty

Z tabulky je patrné, že atribut y_2 nabývá pouze bivalentních logických hodnot, tedy hodnot 0 a 1. Naopak atributy y_1, y_3, y_4 nabývají i jiných hodnot. Jedná se tedy o vícehodnotové kontexty, které můžeme pomocí konceptuálního škálování převést na základní kontext.

Předchozí tabulku tedy s využitím konceptuálního škálování přepíšeme do následující podoby.

	y_1(0-10)	y_1(11-20)	y_1(21-30)	y_2	y_3(0-10)	y_3(11-20)	y_3(21-30)	y_4
x_1	0	1	0	1	1	0	0	0
x_2	1	0	0	1	1	0	0	1
x_3	1	0	0	0	0	0	1	1
x_4	0	0	1	1	0	1	0	0

Tab. 11. Konceptuální škálování

4 INTERNETOVÉ VYHLEDÁVAČE

Dnešní internet se již neobejde bez vyhledávacích serverů neboli vyhledávačů, které nám umožňují najít webové stránky obsahující požadované informace.

Tato kapitola rozděluje internetové vyhledávače do třech základních typů na vyhledávače fulltextové, katalogové a meta vyhledávače a popisuje, jak tyto jednotlivé typy vyhledávačů fungují.

Pro začátek je důležité říci, jak vůbec takové vyhledávače pracují. Vyhledávací servery jako jsou Google, Seznam a Bing obsahují ve své databázi webových stránek mnoho informací o klíčových slovech, která jsou uvedena na daných webových stránkách. Jelikož takové vyhledávače hledají klíčové slovo v celém obsahu webových stránek, říká se takovým vyhledávačům *fulltextové*. Na Internetu jich existují stovky. Na rozdíl od fulltextových vyhledávačů existují i *katalogové* jako například Yahoo! nebo Najisto – katalog Centra a Atlasu. Takovéto vyhledávače už v dnešní době pomalu ani neexistují, protože skoro ve všech případech již jen doplňují fulltextové vyhledávání. Někdy se jim též říká *hybridní* nebo *smíšené* vyhledávače.

4.1 Fulltextové vyhledávače

Fulltextové vyhledávače neboli indexéry se prezentují automatickým procházením webových stránek a pochopením jejich obsahu, k čemuž využívají desítky až miliony počítačů. Prochází webové stránky a stahuje je v určitých intervalech na základě určených algoritmů. Kvalitu procházených a stahovaných stránek vyhodnotí na základě mnoha faktorů. U nejméně významných stahuje jen první stránku v dlouhých intervalech. Při zpracování stažených webových stránek následuje snaha o porozumění textu a jeho tzv. *indexování.* Z toho vyplývá, že je prakticky nemožné, aby do svého výsledku vyhledávání zahrnul text té webové stránky, u které vyhledávač z nějakých důvodů její text nestáhl nebo nezaindexoval. Naopak ve výsledku vyhledávání se objeví v případě, kdy vyhledávač text webové stránky stáhne a zapíše do své databáze nebo ji zaindexuje. Po zapsání webové stránky do databáze následuje třídění neboli pořadí, na kterém místě se může zobrazit. Toto pořadí lze ovlivnit jejich důležitostí, která se ověřuje praktickými testy u každých vyhledávačů zvlášť. Velká množství kritérií, která ovlivňují pořadí, ale nejsou zveřejňována [9].

Kvalita fulltextových vyhledávačů je tedy závislá v prvé řadě na kvalitě výsledných odpovědí, kterou hodnotí uživatel, a to tím jestli jím hledanou informaci najde v prvních místech výsledných odpovědí fulltextového vyhledávače. V databázi vyhledávačů měří kvalitu webových stránek tzv. *Page Rank* (Google), *S-Rank* (Seznam) a *Jyxo Rank* (Jyxo). Majitelé webových stránek se snaží dosáhnout na co nejvyšší pozice ve výstupu vyhledávače. Výsledkem je, že vyhledávač musí své metody neustále vylepšovat, aby vyhověl čím dál vyšším požadavkům svých návštěvníků.

Technologie fulltextových vyhledávačů pracuje ve třech krocích – v automatickém procházení Internetu tzv. *vyhledávacím robotem*, indexováním a poskytováním odpovědí na dotazy [10].

1. **Vyhledávací robot**

Pro procházení webových stránek a vytvoření databáze výskytu slov má fulltextový vyhledávač automatický program na sbírání dat (vyhledávací robot), který se pomocí hypertextových odkazů snaží navštívit všechny webové stránky na Internetu (celý World Wide Web).

Vyhledávač přidělí vyhledávacímu robotovi seznam URL odkazů, nebo seznam rozcestníků jako je Yahoo! nebo katalog Seznamu. Nalezenou stránku si stáhne k sobě na pevný disk a uloží si i její URL adresu pro zaindexování. Zaindexování URL adresy provádí proto, aby webovou stránku nenavštěvoval opakovaně. Další webové stránky nachází na jím stáhnutých webových stránkách, protože se zde většinou nacházejí i hypertextové odkazy na další webové stránky. Tyto webové stránky opět navštíví stejným způsobem. Jelikož webové stránky prochází častými aktualizacemi, pracuje vyhledávací robot cyklicky, aby zjistil jejich případné změny.

Vyhledávací roboti tedy odhalí jen takovou webovou stránku, na kterou vede nějaký odkaz. Odkazů jsou ale řádově miliony, a jelikož není čas projít všechny, tak proto sledují odkazy jen do určitého počtu nebo úrovně [11].

Jak již bylo výše uvedeno, webové stránky, které vyhledávací robot stáhl na svůj pevný disk, musí vyhledávač zpracovat a vytvořit databázi. Tato databáze nabídne všechna vyhledávaná slova a k nim i URL odkaz. Z toho vyplývá, že poskytuje informace, na kterých webových stránkách se vyhledávané slovo vyskytuje. Problémem se může stát velikost databáze, kde by trvalo velmi dlouho její prohledávání, proto existuje indexování.

2. Indexování

V databázi, která je indexovaná, lze velmi rychle najít požadované informace. Index databáze zobrazuje na prvních místech ty webové odkazy, které mají tzv. nejvyšší *relevanci* neboli důležitost v hodnocení kvality. Každý vyhledávač ale nasbírané informace vyhodnotí a zobrazí po svém. Pro určení relevance se využívají algoritmy, které jsou založeny na nejrůznějších znacích stránek a různých úhlech analýzy jejich obsahu.

Důležitost slov

Existuje několik možností, jak lze zvýšit hodnocení webové stránky. Jednou z nich je, když má vyhledávané slovo na stránce vysokou důležitost. Je-li vyhledávané slovo v titulku stránky, nadpisu nebo blíže k začátku stránky, případně se na stránce opakuje, zvyšuje se tím jeho důležitost. Naopak snížením důležitosti se vyhledávač brání, když se vyhledávané slovo nelogicky opakuje, anebo se vloží do obsahu webové stránky bez ohledu na to, jestli jde o skutečný obsah [12].

Klíčová slova

V minulosti byl hlavní důraz kladen na popis a klíčová slova, která měla být seznamem slov charakteristických pro danou webovou stránku. Třeba Google a Seznam klíčová slova úplně ignorují a nepřikládají jim žádnou důležitost.

Popis

Naopak popis webové stránky berou všechny důležité vyhledávače v potaz. Například Google a Seznam zobrazují popis pod výpisem vyhledávání [11].

Atraktivita webové stránky a její serióznost

Jestliže webová stránka odkazuje na více jiných webových stránek, má tato webová stránka vyšší hodnocení z hlediska důležitosti, neboť odkazující webové stránky zřejmě obsahují zajímavé informace. Zvýhodněny jsou ty, které obsahují velké množství kvalitních webových stránek.

Sponzorované odkazy

Jsou jedním ze zdrojů příjmů vyhledávačů. Zpravidla bývají odděleny od neplacených na základě komerčního zvýhodnění. Většina seriózních vyhledávačů se však tomuto zpravidla vyhýbá.

Vyhledávač sleduje	**Důležitost**
titulek	vysoká
klíčové slovo	větší, u Googlu a Seznamu žádná
popis	různá
nadpis	značná
začátek stránky	větší
URL adresa	různá
odkaz z jiného serveru	u Googlu a Jyxo vysoká
text	malá

Tab. 12. Porovnání nejčastějších kritérií z hlediska důležitosti

V tabulce jsou uvedena nejčastější kritéria pro počítání relevance, která se u každého vyhledávače v něčem velmi liší.

3. Poskytování odpovědi na dotazy

Každý vyhledávač třídí a zobrazuje výsledky na dotazy různě. Dá se říci, čím více se hledaná slova na stránce vyskytují a mají důležitější pozici (titulek a nadpis), tím je webová stránka zobrazena ve výsledku výše. Algoritmy, které zobrazují výsledky vyhledávání, nikdo nezná a liší se u každého vyhledávače.

SEO *(Search Engine Optimalization)*

Velmi žádanou službou v poslední době je technika SEO, která dokáže webové stránky upravit tak, aby se umístily co nejvýše ve výsledcích vyhledávání jednotlivých vyhledávačů. Z hlediska vyhledávačů je však jakékoli umělé zlepšování kvality webových stránek ve výsledcích vyhledávání „zakázané“. Tzv. *Black Hat SEO* vyhledávače postihují vyřazením webových stránek ze svého indexu [12].

Page Rank

Google ale používá i další metodu, která vyjadřuje důležitost webové stránky a ovlivňuje třídění a řazení výsledků vyhledávání. Je jím tzv. *Page Rank*. U každé webové stránky se dá změřit Page Rank pomocí jednoduchého programu od Googlu. Program počítá s tím, kolik webových stránek se na danou webovou stránku odkazuje (čím vyšší

Page Rank odkazových webových stránek, tím vyšší důležitost). Tedy čím více odkazů, tím lepší je i Page Rank. Důležitá je také četnost aktualizací a mnoho dalších faktorů.

Nejlepší český vyhledávač Seznam má tzv. *S-Rank* a Jyxo tzv. *Jyxo Rank*, který se počítá podle toho, z kolika různých domén druhé úrovně vedou na webovou stránku odkazy [11].

Mezi nejznámější české fulltextové vyhledávače patří Seznam s podílem vyhledávání přibližně 52% a s velkým odstupem Jyxo. Ve světě patří mezi nejvýznamnější a největší fulltextové vyhledávače Google přes 90% a s velkým odstupem Bing.

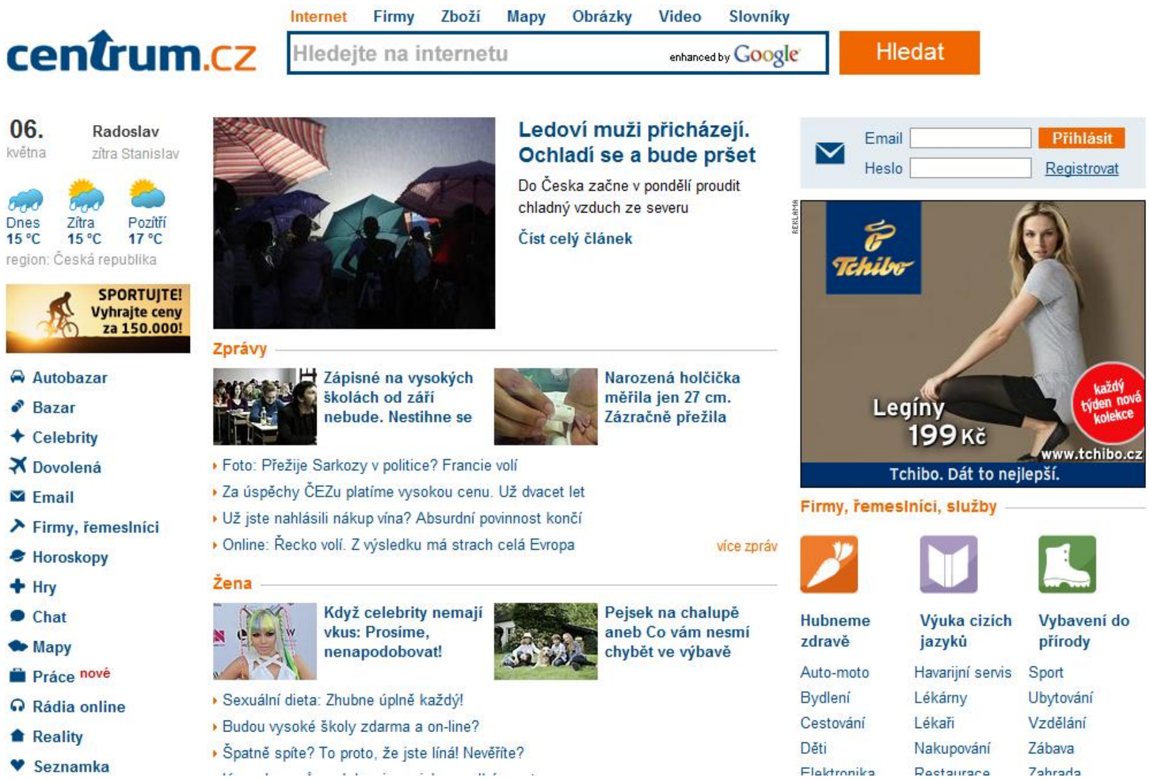
4.1.1 Seznam

Seznam byl první katalogový vyhledávač a posléze jako první v ČR začal využívat vlastní fulltextovou technologii pro vyhledávání v ČR (vlastní reklamy Sklik) a fulltextovou technologii Bing od Microsoftu pro vyhledávání ve světě. Vedle fulltextového vyhledávače obsahuje i katalog, který je rozdělen na komerční a nekomerční část. Uživatelům se nabízí snadnější a rychlejší přístup ke strukturované informaci [13], [14].

Obr. 16. Rozhraní vyhledávače Seznam

4.1.2 Centrum

Centrum Holdings a.s. provozuje portál Centrum i Atlas od roku 2008. Portál Centrum nemá vlastní fulltextové vyhledávání a je tedy zajišťován Googlem, který dodává i reklamy pomocí PPC systému AdWords [15].



Obr. 17. Rozhraní vyhledávače Centrum

4.1.3 Jyxo

Pro vyhledávání rozsáhlého množství dat využívá Jyxo svou vlastní technologii, která má modulární strukturu a podporuje tak čtení a vyhledávání ve více formátech (html, pdf nebo doc). Umožňuje prohledání databází a vyhledání audiovizuálních dat a obrázků. Pro svoje vlastní hodnocení stránek využívá tzv. *Jyxo Rank* [16], [17].

Obr. 18. Rozhraní vyhledávače Jyxo

4.1.4 Google

Díky rychlosti vyhledávání, vysoké relevanci výsledků a jednoduchému vzhledu je Google v současnosti nejpoužívanější internetový vyhledávač na světě.

Kvalitu webových stránek určuje díky Page Rank, který byl vysvětlen již výše. Google ve výsledcích vyhledávání zobrazí ve vrchní části nejprve placené PPC odkazy, tzv. *Google Adwords*, a poté přirozené výsledky. Placeným odkazům patří ještě pravá část a spodní strana stránky.

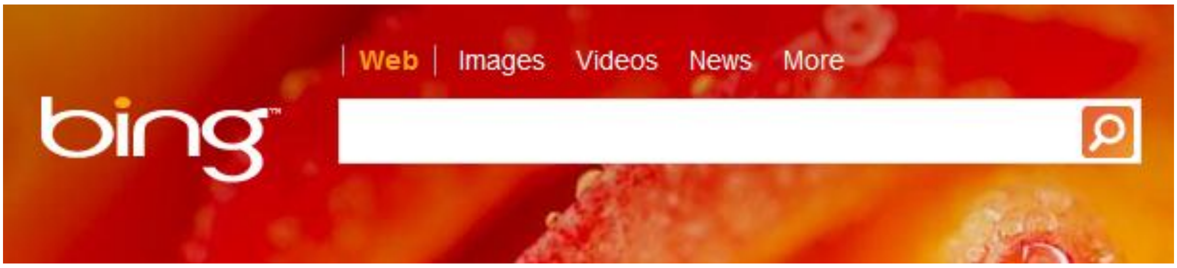
Mezi další vyhledávače, které využívají jeho technologii, patří AOL, Earthlink, Centrum a řada dalších. Celkový podíl uživatelů používající Google, přesahuje 90% [18], [19], [21].

Obr. 19. Rozhraní vyhledávače Google

4.1.5 Bing

Bing (dřívější název Live Search, Windows Live Search a MSN Search) je internetový vyhledávač používaný společností Microsoft. Spolupracuje s vyhledávačem Yahoo!, kterému zajišťuje organické vyhledávání a sdílí s ním reklamní systém Microsoft adCenter.

I když Bing patří mezi nejpoužívanější internetové vyhledávače, jeho podíl na celosvětovém trhu vyhledávačů je velmi malý a pouze v USA překračuje hranici 10% [20].



Obr. 20. Rozhraní vyhledávače Bing

4.2 Katalogové vyhledávače

Katalogové vyhledávače jsou projekty velkých portálů, jako je Seznam, Centrum nebo Yahoo!. Obsahují velkou databázi odkazů na různé webové stránky celého Internetu, uspořádanou do stromové struktury a řazenou do jednotlivých kategorií a podkategorií podle tematických sekcí (např. zábava, vzdělání, kultura nebo sport). Jestliže chceme přidat webovou stránku do katalogu, musíme ji zaregistrovat na příslušné registrační stránce, kde je poté fyzicky zkontrolována a schvalována lidským editorem (pracovníkem katalogu), který posuzuje také její vhodnost pro danou kategorii. Teprve potom je webová stránka do katalogu přidána. Při registraci do katalogu je třeba zadat URL adresu webové stránky, titulek, který bude v katalogu sloužit současně jako odkaz na webovou stránku, doplňující popis webové stránky a zvolit vhodnou kategorii pro registraci [10].

Existují dva způsoby, jak hledat v katalogovém vyhledávači. Jeden způsob je procházení databáze zaregistrovaných webových stránek a druhý způsob je procházení přímo v kategoriích vyhledávače. Výsledek vyhledávání zobrazuje katalogový vyhledávač jako ručně vkládané a editované popisy webových stránek [22].

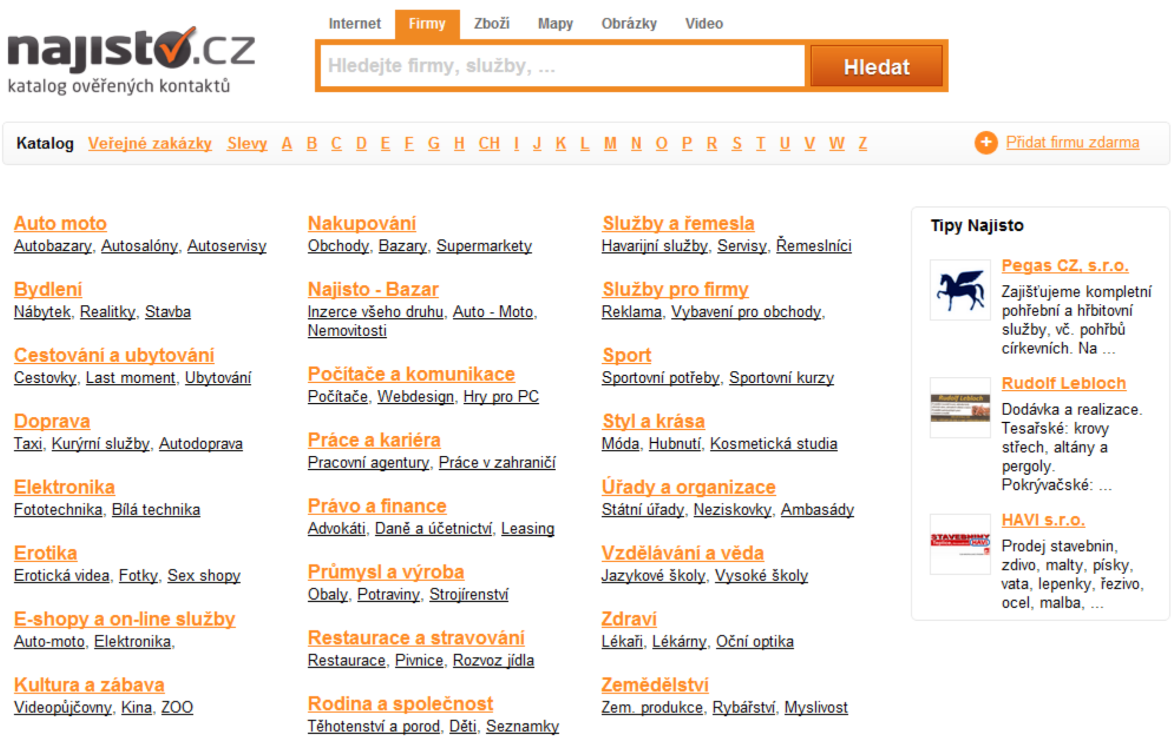
Pozice webových stránek ovlivní pouze zařazení do správné (relevantní) kategorie, anebo umístění klíčových slov do titulku odkazu a popisu. Webová stránka se zobrazí uživatelům, kteří hledají daná klíčová slova. Výhodu mají webové stránky, které obsahují klíčové slovo přímo ve svém názvu nebo v URL adrese webové stránky (například lednicky.cz) [10].

Centrum a Atlas už svoje katalogové vyhledávače nemají, ale místo toho provozují katalog Najisto, což vlastně už vůbec není katalog, ale pokus o prodej firemních zápisů [23].

Mezi nejpoužívanější katalogové vyhledávače u nás patří katalogy tří největších českých vyhledávacích portálů (Seznam a jeho firemní katalog Firmy a katalog Najisto, který je společný pro portály Centrum a Atlas [10].

4.2.1 Najisto

Jedná se o druhý největší a nejnavštěvovanější katalogový vyhledávač firem v ČR, hned po firmy.cz od Seznamu. Nabízí provázanost a přednostní výpis s vyhledávači Centrum a Atlas.

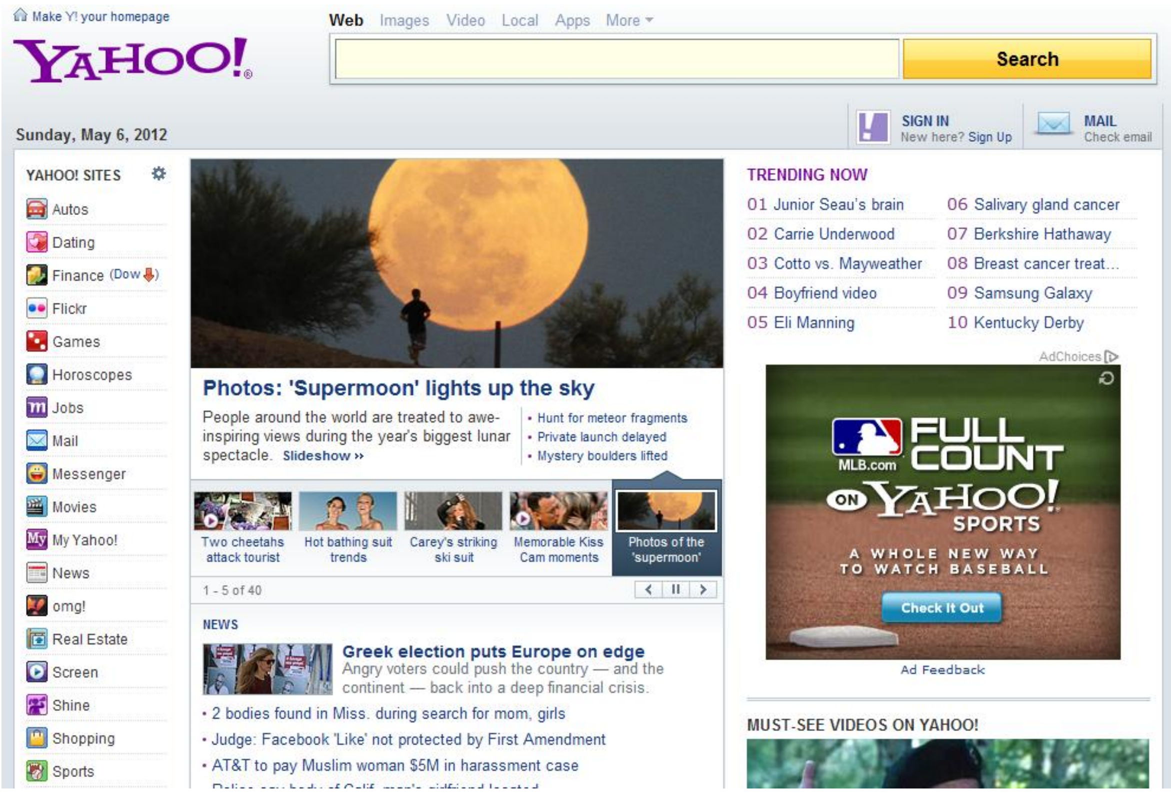


Obr. 21. Rozhraní vyhledávače Najisto

4.2.2 Yahoo!

Yahoo! kdysi býval velmi populárním katalogovým vyhledávačem. V současnosti je relativně používaný v USA, v celosvětovém měřítku vyhledávačů se pohybuje okolo 4%. Vedle katalogu má i fulltextové vyhledávání, které obstarává internetový vyhledávač Bing společnosti Microsoft.

Při vyhledávání zobrazuje na prvních místech placené odkazy systému Yahoo! Search Marketing a až pod nimi přirozené výsledky. V roce 2010 se spojil s reklamním systémem Microsoft adCenter. Placené odkazy zobrazuje podobně jako Google na pravé straně [24], [25].

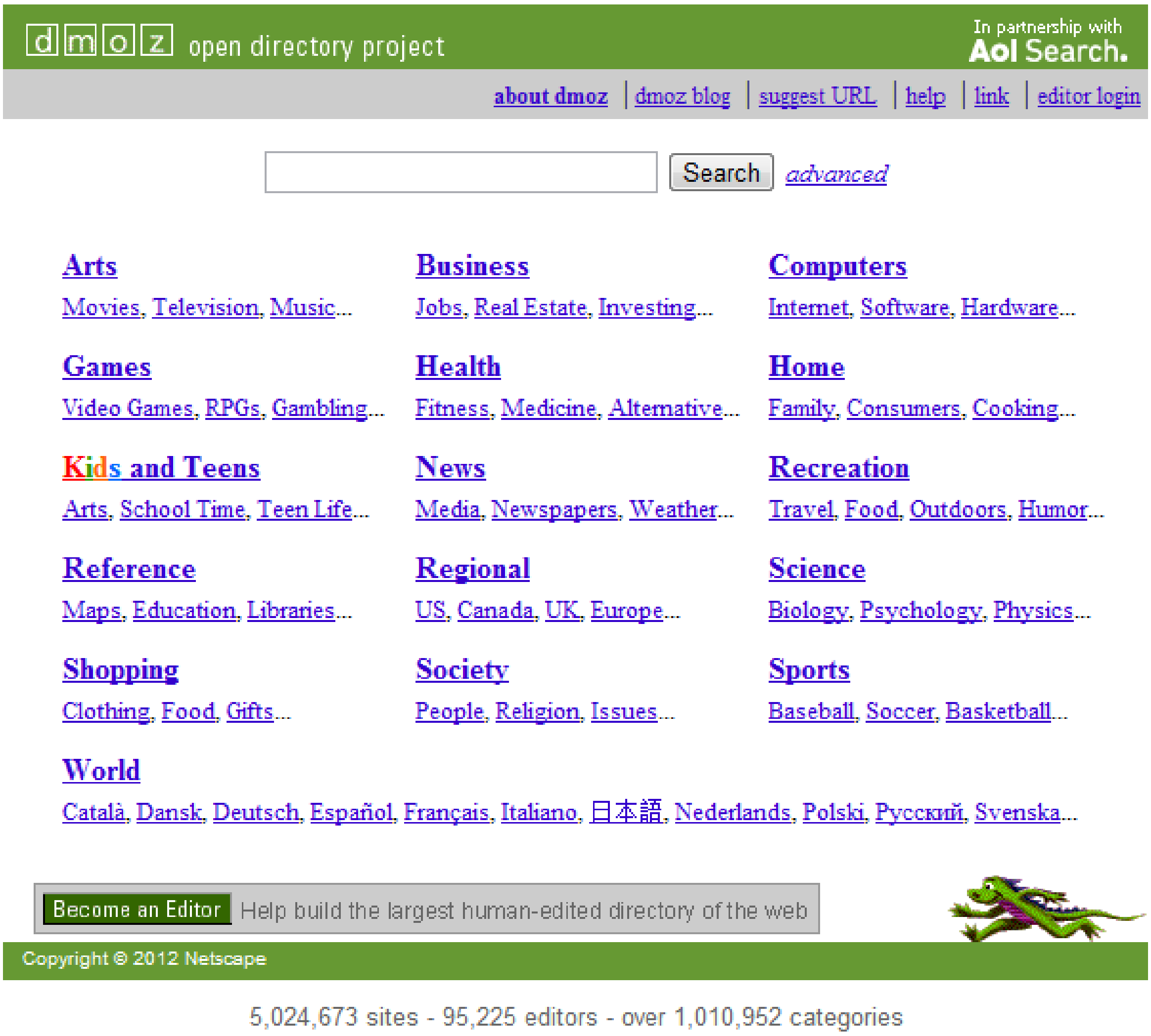


Obr. 22. Rozhraní vyhledávače Yahoo!

4.2.3 DMOZ

Open Directory Project neboli DMOZ je největší lidmi budovaný katalog webových stránek, který sestavují dobrovolní editoři z celého světa. Na základě toho musí být pro jednotlivé odkazy rozdělen do několika jazykových sekcí.

Jelikož u katalogu DMOZ nejsou žádné webové stránky zvýhodněny, nesetkáme se tak u něho s placenými odkazy. Cílem není vytvořit katalog všech webových stránek na Internetu, ale jen jejich kvalitní a informačně hodnotný výběr [26].

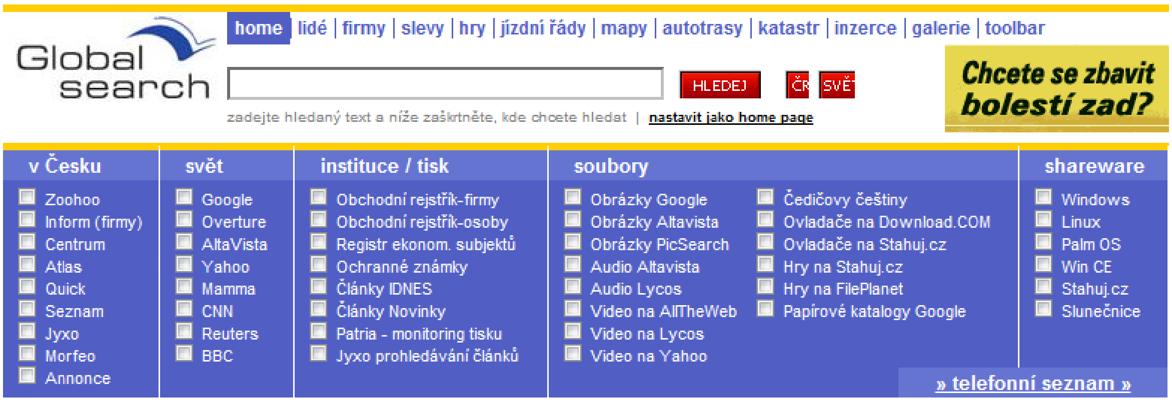


Obr. 23. Rozhraní vyhledávače DMOZ

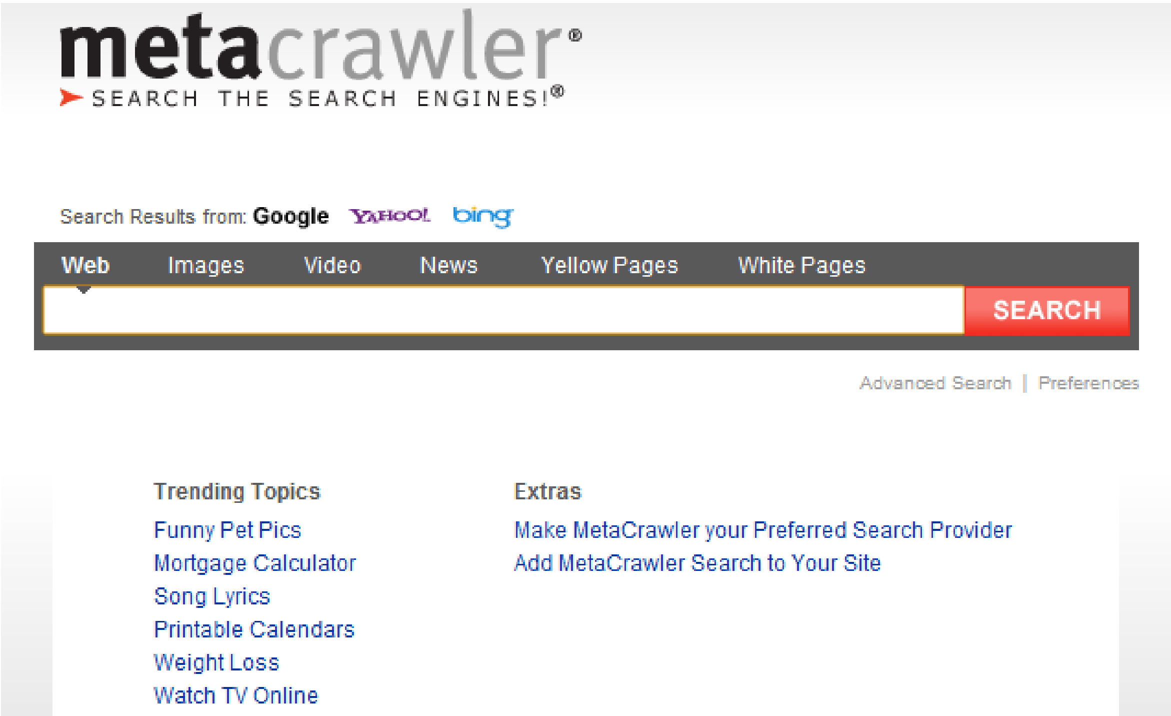
4.3 Meta vyhledávače

Meta vyhledávače jsou takové vyhledávače, které nemají vlastní vyhledávací stroj, tudíž ani vlastní databázi webových stránek, ale používají a spoléhají se při zadání dotazu na jiné vyhledávače. Takto odeslaný dotaz poté vyhodnotí a zpracují výsledek.

Představitelé meta vyhledávačů jsou například Globalsearch v ČR a Metacrawler v zahraničí [27].



Obr. 24. Rozhraní vyhledávače Globalsearch



Obr. 25. Rozhraní vyhledávače Metacrawler

5 APLIKACE FCA

Metodou FCA můžeme rozšířit možnosti internetových vyhledávačů, a tím odstranit jejich současné hlavní nedostatky, které aplikují pro uživatelský dotaz statické metody, jako jsou například četnost výskytu slov nebo histogram. FCA zobrazí relevantní výsledky na uživatelský dotaz tím, že vytvoří větší konceptuální svaz z obsáhlejšího kontextu, který internetové vyhledávače dostaly z uživatelského dotazu.

5.1 Srovnání vyhledávačů

5.1.1 České vyhledávače

Na šest českých internetových vyhledávačů a jejich atributy použijeme FCA. Jako atributy byly zvoleny unikátní návštěvníci (návštěvníci, kteří navštíví webovou stránku za měsíc – při opakované návštěvě stejným návštěvníkem se jeho návštěva již nezapočítává), typ vyhledávače, vyhledání dokumentů (pdf, doc, ppt), více jazyků, třídění výsledků, náhled webové stránky a vyhledání videa a obrázků. [28], [29], [30], [31]

Množina objektů:

$X = \{seznam.cz, centrum.cz, najisto.cz, jyxo.cz, caramba.cz, globalsearch.cz\}$

Množina atributů:

$Y = \{Unikátní\ návštěvníci, Typ, Vyhledávání\ dokumentů, Více\ jazyků, Třídění\ výsledků, Náhled\ webové\ stránky, Vyhledávání\ videa\ a\ obrázků\}$

G	**Unikátní návštěvníci**	**Typ**	**Vyhledání dokumentů pdf,doc,ppt**	**Více jazyků**	**Třídění výsledků**	**Náhled webové stránky**	**Vyhledání videa a obrázků**
seznam.cz	5 528 197	fulltextový	ano	ne	ano	ano	ano
centrum.cz	3 136 642	fulltextový	ne	ne	ano	ne	ano
najisto.cz	613 809	katalogový	ne	ne	ano	ne	ne
jyxo.cz	55 298	fulltextový	ano	ne	ano	ne	ne
caramba.cz	15 778	katalogový	ne	ne	ano	ne	ne
globalsearch.cz	14 736	meta					

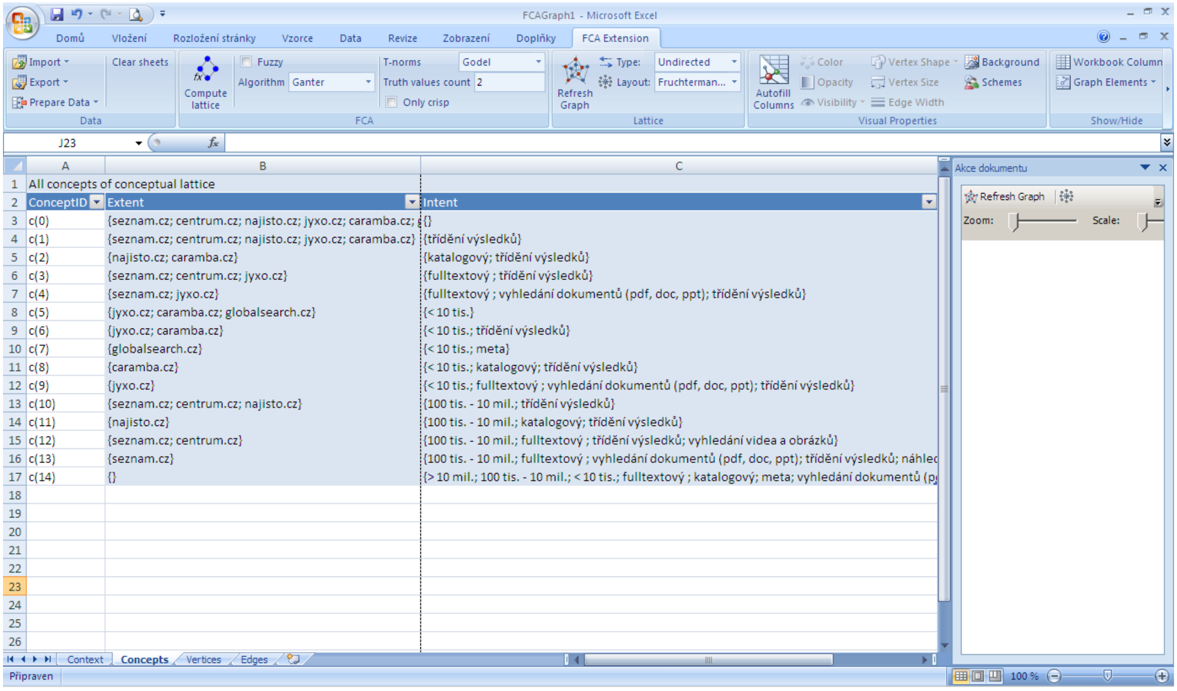
Tab. 13. Formální kontext českých internetových vyhledávačů

Na předchozí tabulku použijeme konceptuální škálování, které převede hodnoty tabulky do bivalentních hodnot.

G	Unikátní návštěvníci			Typ			Vyhledání dokumentů pdf,doc,ppt	Více jazyků	Třídění výsledků	Náhled webové stránky	Vyhledání videa a obrázků
	> 10 mil.	100 tis. - 10 mil.	< 100 tis.	fulltextový	katalogový	meta					
seznam.cz	0	1	0	1	0	0	1	0	1	1	1
centrum.cz	0	1	0	1	0	0	0	0	1	0	1
najisto.cz	0	1	0	0	1	0	0	0	1	0	0
jyxo.cz	0	0	1	1	0	0	1	0	1	0	0
caramba.cz	0	0	1	0	1	0	0	0	1	0	0
globalsearch.cz	0	0	1	0	0	1					

Tab. 14. Formální kontext českých internetových vyhledávačů po procesu škálování

Z bivalentních hodnot tabulky odvodíme jednotlivé formální koncepty pomocí programu FCA Extension.



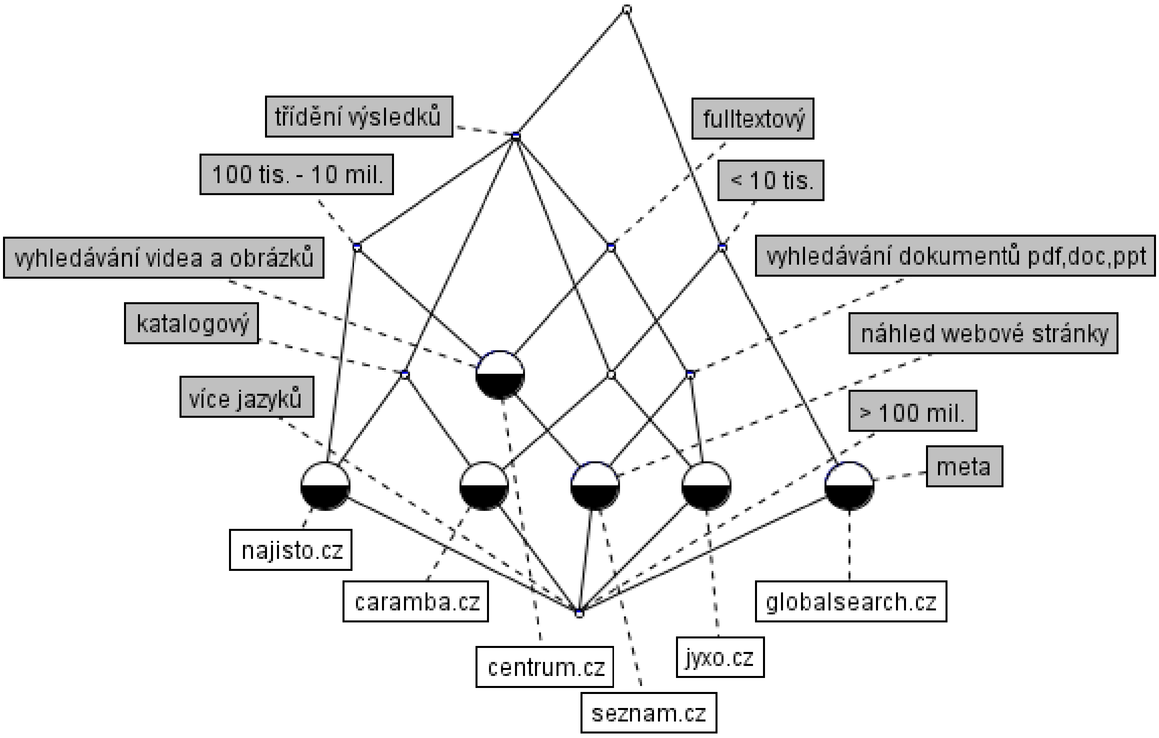
Obr. 26. Rozhraní programu FCA Extension v MS Excel

Seznam konceptů:

{seznam.cz; centrum.cz; najisto.cz; jyxo.cz; caramba.cz; globalsearch.cz}	{}
{seznam.cz; centrum.cz; najisto.cz; jyxo.cz; caramba.cz}	{třídění výsledků}
{najisto.cz; caramba.cz}	{katalogový; třídění výsledků}
{seznam.cz; centrum.cz; jyxo.cz}	{fulltextový; třídění výsledků}
{seznam.cz; jyxo.cz}	{fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků}
{jyxo.cz; caramba.cz; globalsearch.cz}	{< 10 tis.}
{jyxo.cz; caramba.cz}	{< 10 tis.; třídění výsledků}
{globalsearch.cz}	{< 10 tis.; meta}
{caramba.cz}	{< 10 tis.; katalogový; třídění výsledků}
{jyxo.cz}	{< 10 tis.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků}
{seznam.cz; centrum.cz; najisto.cz}	{100 tis. - 10 mil.; třídění výsledků}
{najisto.cz}	{100 tis. - 10 mil.; katalogový; třídění výsledků}
{seznam.cz; centrum.cz}	{100 tis. - 10 mil.; fulltextový; třídění výsledků; vyhledání videa a obrázků}
{seznam.cz}	{100 tis. - 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků; náhled webové stránky; vyhledání videa a obrázků}
{}	{> 10 mil.; 100 tis. - 10 mil.; < 10 tis.; fulltextový; katalogový; meta; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; náhled webové stránky; vyhledání videa a obrázků}

Tab. 15. Množina formálních konceptů českých internetových vyhledávačů

Konceptuální svaz:



Obr. 27. Konceptuální svaz českých internetových vyhledávačů

5.1.2 Zahraniční vyhledávače

V tomto příkladu použijeme FCA na sedm zahraničních internetových vyhledávačů a jejich atributy. Atributy byly zvoleny stejné jako z předchozího příkladu, tudíž unikátní návštěvníci, typ vyhledávače, vyhledání dokumentů (pdf, doc, ppt), více jazyků, třídění výsledků, náhled webové stránky a vyhledání videa a obrázků [28], [29], [30], [31].

Množina objektů:

$X = \{google.com, yahoo.com, bing.com, ask.com, altavista.com, dmoz.org,$

$metacrawler.com\}$

Množina atributů:

$Y = \{Unikátní\ návštěvníci, Typ, Vyhledávání\ dokumentů, Více\ jazyků, Třídění$

$výsledků, Náhled\ webové\ stránky, Vyhledávání\ videa\ a\ obrázků\}$

G	Unikátní návštěvníci	Typ	Vyhledání dokumentů pdf,doc,ppt	Více jazyků	Třídění výsledků	Náhled webové stránky	Vyhledání videa a obrázků
google.com	176 255 305	fulltextový	ano	ano	ano	ne	ano
yahoo.com	154 680 428	katalogový	ano	ano	ano	ne	ano
bing.com	105 687 931	fulltextový	ne	ano	ano	ne	ne
ask.com	72 059 717	fulltextový	ne	ne	ano	ne	ano
altavista.com	304 174	fulltextový	ano	ano	ano	ne	ano
dmoz.org	264 968	katalogový	ne	ano	ano	ne	ne
metacrawler.com	96 688	meta					

Tab. 16. Formální kontext zahraničních internetových vyhledávačů

Opět na předchozí tabulku použijeme konceptuální škálování, které nám převede hodnoty tabulky na bivalentní.

G	Unikátní návštěvníci			Typ			Vyhledání dokumentů pdf,doc,ppt	Více jazyků	Třídění výsledků	Náhled webové stránky	Vyhledání videa a obrázků
	> 10 mil.	100 tis. - 10 mil.	< 100 tis.	fulltextový	katalogový	meta					
google.com	1	0	0	1	0	0	1	1	1	0	1
yahoo.com	1	0	0	0	1	0	1	1	1	0	1
bing.com	1	0	0	1	0	0	0	1	1	0	0
ask.com	1	0	0	1	0	0	0	0	1	0	1
altavista.com	0	1	0	1	0	0	1	1	1	0	1
dmoz.org	0	1	0	0	1	0	0	1	1	0	0
metacrawler.com	0	0	1	0	0	1					

Tab. 17. Formální kontext zahraničních internetových vyhledávačů po procesu škálování

Z bivalentních hodnot tabulky opět odvodíme jednotlivé formální koncepty pomocí programu FCA Extension. Jejich množina bude vypadat následovně.

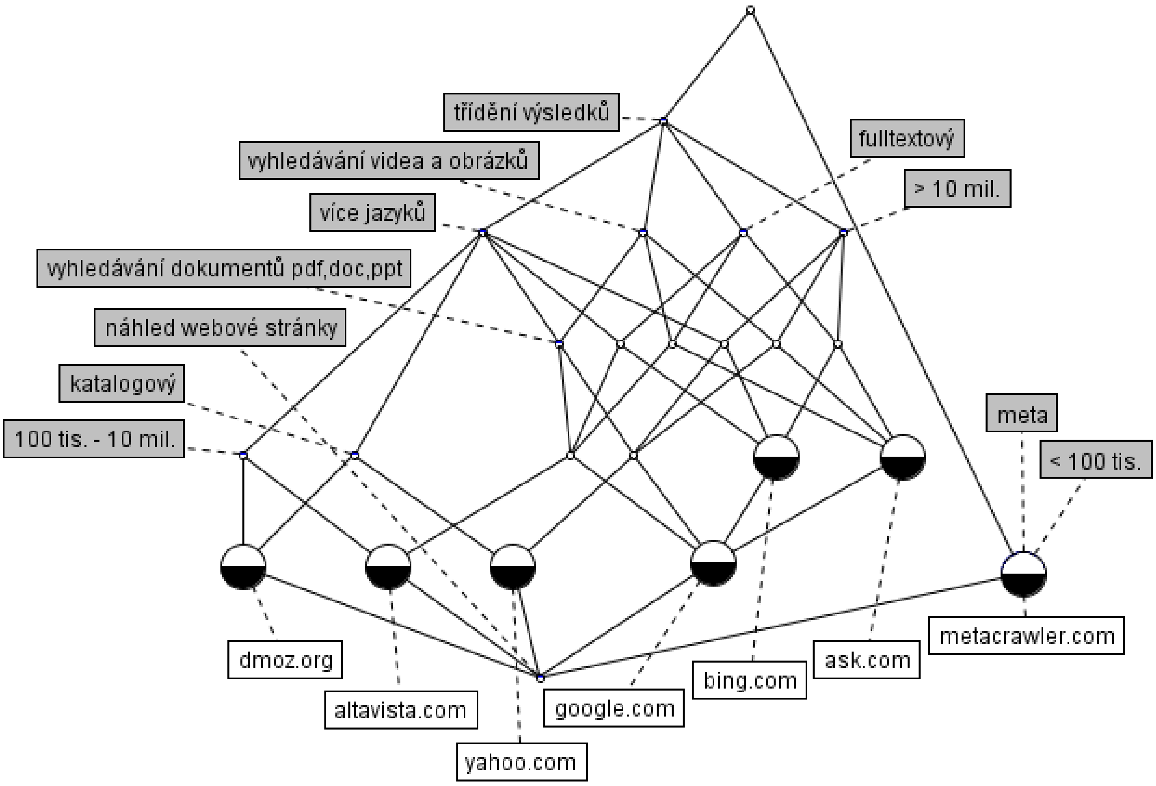
Seznam konceptů:

{google.com; yahoo.com; bing.com; ask.com; altavista.com; dmoz.org; metacrawler.com}	{}
{google.com; yahoo.com; bing.com; ask.com; altavista.com; dmoz.org}	{třídění výsledků}
{google.com; yahoo.com; ask.com; altavista.com}	{třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; bing.com; altavista.com; dmoz.org}	{více jazyků; třídění výsledků}
{google.com; yahoo.com; altavista.com}	{vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{yahoo.com; dmoz.org}	{katalogový; více jazyků; třídění výsledků}
{google.com; bing.com; ask.com; altavista.com}	{fulltextový; třídění výsledků}
{google.com; ask.com; altavista.com}	{fulltextový; třídění výsledků; vyhledání videa a obrázků}
{google.com; bing.com; altavista.com}	{fulltextový; více jazyků; třídění výsledků}
{google.com; altavista.com}	{fulltextový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{metacrawler.com}	{< 10 tis.; meta}
{altavista.com; dmoz.org}	{100 tis. - 10 mil.; více jazyků; třídění výsledků}
{dmoz.org}	{100 tis. - 10 mil.; katalogový; více jazyků; třídění výsledků}
{altavista.com}	{100 tis. - 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; bing.com; ask.com}	{> 10 mil.; třídění výsledků}
{google.com; yahoo.com; ask.com}	{> 10 mil.; třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; bing.com}	{> 10 mil.; více jazyků; třídění výsledků}
{google.com; yahoo.com}	{> 10 mil.; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{yahoo.com}	{> 10 mil.; katalogový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{google.com; bing.com; ask.com}	{> 10 mil.; fulltextový; třídění výsledků}
{google.com; ask.com}	{> 10 mil.; fulltextový; třídění výsledků; vyhledání videa a obrázků}
{google.com; bing.com}	{> 10 mil.; fulltextový; více jazyků; třídění výsledků}
{google.com}	{> 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}

{}	{> 10 mil.; 100 tis. - 10 mil.; < 10 tis.; fulltextový; katalogový; meta; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; náhled webové stránky; vyhledání videa a obrázků}

Tab. 18. Množina formálních konceptů zahraničních internetových vyhledávačů

Konceptuální svaz:



Obr. 28. Konceptuální svaz zahraničních internetových vyhledávačů

5.1.3 České a zahraniční vyhledávače

Jako poslední příklad použijeme porovnání českých a zahraničních internetových vyhledávačů. Pro zachování objektivnosti zůstávají atributy neměnné, stejně jako v předchozích příkladech.

Množina objektů:

$X = \{seznam.cz, centrum.cz, najisto.cz, jyxo.cz, caramba.cz, globalsearch.cz,$

$google.com, yahoo.com, bing.com, ask.com, altavista.com, dmoz.org,$

$metacrawler.com\}$

Množina atributů:

$Y = \{Unikátní\ návštěvníci, Typ, Vyhledávání\ dokumentů, Více\ jazyků, Třídění\ výsledků, Náhled\ webové\ stránky, Vyhledávání\ videa\ a\ obrázků\}$

Je zde vynechána tabulka, která obsahovala stejné hodnoty z předchozích příkladů, které byly věnovány zvlášť českým a zahraničním internetovým vyhledávačům.

G	Unikátní návštěvníci			Typ			Vyhledání dokumentů pdf,doc,ppt	Více jazyků	Třídění výsledků	Náhled webové stránky	Vyhledání videa a obrázků
	> 10 mil.	100 tis. - 10 mil.	< 100 tis.	fulltextový	katalogový	meta					
seznam.cz	0	1	0	1	0	0	1	0	1	1	1
centrum.cz	0	1	0	1	0	0	0	0	1	0	1
najisto.cz	0	1	0	0	1	0	0	0	1	0	0
jyxo.cz	0	0	1	1	0	0	1	0	1	0	0
caramba.cz	0	0	1	0	1	0	0	0	1	0	0
globalsearch.cz	0	0	1	0	0	1					
google.com	1	0	0	1	0	0	1	1	1	0	1
yahoo.com	1	0	0	0	1	0	1	1	1	0	1
bing.com	1	0	0	1	0	0	0	1	1	0	0
ask.com	1	0	0	1	0	0	0	0	1	0	1
altavista.com	0	1	0	1	0	0	1	1	1	0	1
dmoz.org	0	1	0	0	1	0	0	1	1	0	0
metacrawler.com	0	0	1	0	0	1					

Tab. 19. Formální kontext českých a zahraničních internetových vyhledávačů po procesu škálování

Z bivalentních hodnot tabulky odvodíme jednotlivé formální koncepty pomocí programu FCA Extension. Jejich množina bude vypadat následovně.

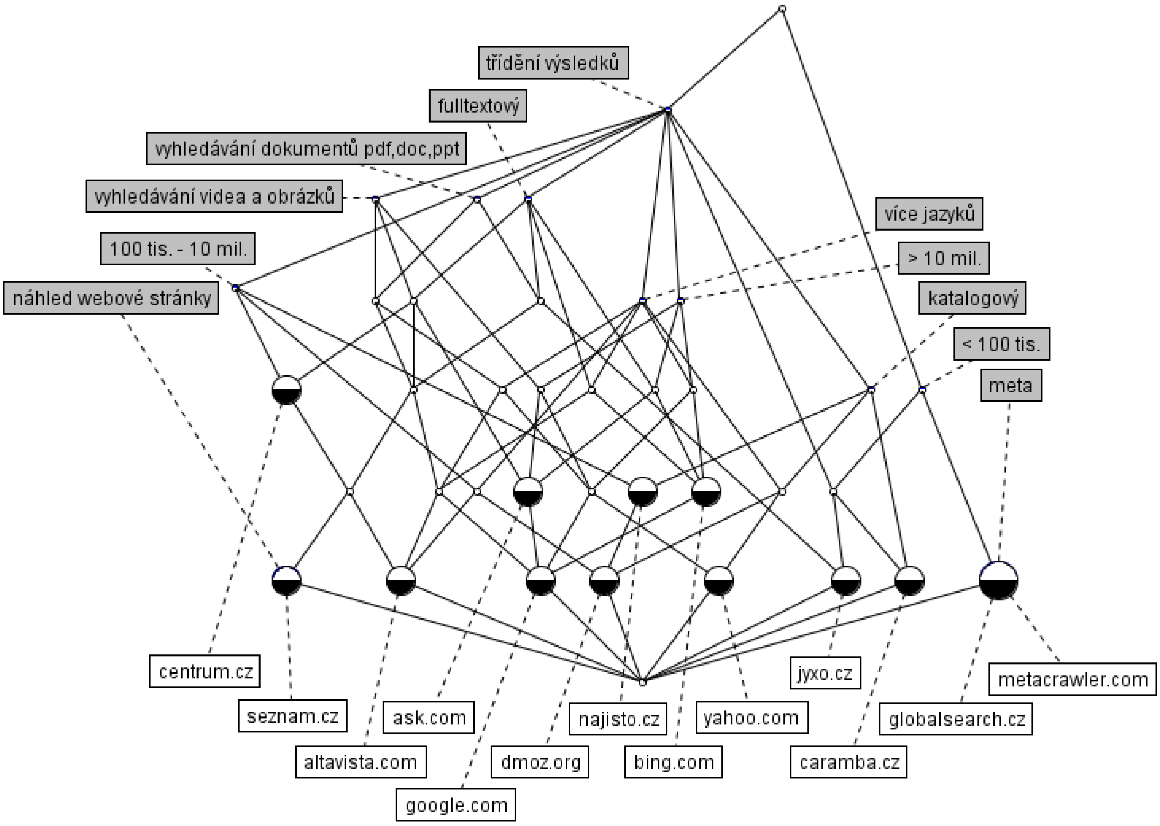
Seznam konceptů:

{seznam.cz; centrum.cz; najisto.cz; jyxo.cz; caramba.cz; globalsearch.cz; google.com; yahoo.com; bing.com; ask.com; altavista.com; dmoz.org; metacrawler.com}	{}
{seznam.cz; centrum.cz; najisto.cz; jyxo.cz; caramba.cz; google.com; yahoo.com; bing.com; ask.com; altavista.com; dmoz.org}	{třídění výsledků}
{seznam.cz; centrum.cz; google.com; yahoo.com; ask.com; altavista.com}	{třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; bing.com; altavista.com; dmoz.org}	{více jazyků; třídění výsledků}
{seznam.cz; jyxo.cz; google.com; yahoo.com; altavista.com}	{vyhledání dokumentů (pdf, doc, ppt); třídění výsledků}
{seznam.cz; google.com; yahoo.com; altavista.com}	{vyhledání dokumentů (pdf, doc, ppt); třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; altavista.com}	{vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{najisto.cz; caramba.cz; yahoo.com; dmoz.org}	{katalogový; třídění výsledků}
{yahoo.com; dmoz.org}	{katalogový; více jazyků; třídění výsledků}
{seznam.cz; centrum.cz; jyxo.cz; google.com; bing.com; ask.com; altavista.com}	{fulltextový; třídění výsledků}
{seznam.cz; centrum.cz; google.com; ask.com; altavista.com}	{fulltextový; třídění výsledků; vyhledání videa a obrázků}
{google.com; bing.com; altavista.com}	{fulltextový; více jazyků; třídění výsledků}
{seznam.cz; jyxo.cz; google.com; altavista.com}	{fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků}
{seznam.cz; google.com; altavista.com}	{fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků; vyhledání videa a obrázků}
{google.com; altavista.com}	{fulltextový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{jyxo.cz; caramba.cz; globalsearch.cz; metacrawler.com}	{< 10 tis.}
{jyxo.cz; caramba.cz}	{< 10 tis.; třídění výsledků}
{globalsearch.cz; metacrawler.com}	{< 10 tis.; meta}
{caramba.cz}	{< 10 tis.; katalogový; třídění výsledků}

{jyxo.cz}	{< 10 tis.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků}
{seznam.cz; centrum.cz; najisto.cz; altavista.com; dmoz.org}	{100 tis. - 10 mil.; třídění výsledků}
{altavista.com; dmoz.org}	{100 tis. - 10 mil.; více jazyků; třídění výsledků}
{najisto.cz; dmoz.org}	{100 tis. - 10 mil.; katalogový; třídění výsledků}
{dmoz.org}	{100 tis. - 10 mil.; katalogový; více jazyků; třídění výsledků}
{seznam.cz; centrum.cz; altavista.com}	{100 tis. - 10 mil.; fulltextový; třídění výsledků; vyhledání videa a obrázků}
{seznam.cz; altavista.com}	{100 tis. - 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků; vyhledání videa a obrázků}
{seznam.cz}	{100 tis. - 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); třídění výsledků; náhled webové stránky; vyhledání videa a obrázků}
{altavista.com}	{100 tis. - 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; bing.com; ask.com}	{> 10 mil.; třídění výsledků}
{google.com; yahoo.com; ask.com}	{> 10 mil.; třídění výsledků; vyhledání videa a obrázků}
{google.com; yahoo.com; bing.com}	{> 10 mil.; více jazyků; třídění výsledků}
{google.com; yahoo.com}	{> 10 mil.; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{yahoo.com}	{> 10 mil.; katalogový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{google.com; bing.com; ask.com}	{> 10 mil.; fulltextový; třídění výsledků}
{google.com; ask.com}	{> 10 mil.; fulltextový; třídění výsledků; vyhledání videa a obrázků}
{google.com; bing.com}	{> 10 mil.; fulltextový; více jazyků; třídění výsledků}
{google.com}	{> 10 mil.; fulltextový; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; vyhledání videa a obrázků}
{}	{> 10 mil.; 100 tis. - 10 mil.; < 10 tis.; fulltextový; katalogový; meta; vyhledání dokumentů (pdf, doc, ppt); více jazyků; třídění výsledků; náhled webové stránky; vyhledání videa a obrázků}

Tab. 20. Množina formálních konceptů českých a zahraničních internetových vyhledávačů

Konceptuální svaz:



Obr. 29. Konceptuální svaz českých a zahraničních internetových vyhledávačů

Při hledání na Internetu volí většina uživatelů fulltextové vyhledávače, protože při správném zadání dotazu mají relativně vysokou šanci, že požadovanou webovou stránku nebo informaci najdou. U katalogových vyhledávačů jsou omezeni webmastrem, u kterého se musejí spolehnout na pečlivé zadání webové stránky, kterou přidal do katalogu. U meta vyhledávačů mají uživatelé problém v tom, že nikdy nevědí, od jakého vyhledávače je výsledný dotaz, popř. jak kvalitní výsledný dotaz je.

Většina například českých portálů dnes využívá vyhledávání kombinovanou formou, tedy jak vyhledávání v katalogu, tak i nějakou fulltextovou technologií často převzatou (např. Centrum – používá katalog Najisto a fulltextové vyhledávání Googlem) [10].

Formální konceptuální analýza je metodou explorativní neboli průzkumové analýzy dat, která se nejčastěji využívá v tzv. *data miningu*. Data mining představuje rozsáhlé odvětví zabývající se získáváním netriviálních skrytých a potencionálně užitečných informací z již existujících naplněných databází informací. Při dobývání znalostí z databází využívá tato informační technologie zejména strojové učení a další obory umělé inteligence. Jestliže použijeme správný způsob získávání informací, lze předpokládat, že ze správných dat dostaneme správné výsledky. Ovšem zda jsme zvolili právě ten nejvhodnější způsob, patří ke klíčovým problémům, stejně tak jako nalezení zajímavého vzorku (clustering), popis vzorku (clasification) a hledání závislostí. Výsledné formální koncepty pak zobrazují zajímavé souvislosti formální konceptuální analýzy, které lze považovat za shluky vyskytující se v objekt – atributových datech, a jsou metodami shlukové analýzy (cluster analysis) dat.

Každý vyhledávač má svůj vlastní algoritmus, pomocí něhož vytváří seznam webových stránek podle relevance vyhledaných řešení na zadaný dotaz. Teprve zde dochází k aplikaci formální konceptuální analýzy, pomocí níž se ze seřazeného seznamu stránek vytváří kontext, a umožňuje tak zefektivnit celý proces vyhledávání dokumentů či webových stránek.

ZÁVĚR

V informatice představuje formální konceptuální analýza způsob odvozování konceptů v rámci určité hierarchie. Každý koncept v této hierarchii reprezentuje množinu objektů, které sdílejí stejné hodnoty pro určitý soubor vlastností, a každá dílčí koncepce v hierarchii obsahuje další podmnožinu objektů ve stejném pojetí jako jí nadřazená množina objektů. Původním smyslem formální konceptuální analýzy byla konkrétní reprezentace úplných svazů a jejich vlastností pomocí datových tabulek, které představují binární vztahy mezi objekty a atributy. Tuto analýzu zveřejnil Rudolf Wille v roce 1984 a navázal tak na práci Birkhoffa a dalších v roce 1930.

Cílem této diplomové práce bylo zpracovat základy formální konceptuální analýzy a formulovat základní reprezentační větu pomocí Galoisových konexí, uvést přehled jednotlivých internetových vyhledávačů, popsat základní principy jejich činnosti a provést rozbor internetových vyhledávačů a poskytovaných služeb právě metodami formální konceptuální analýzy.

Práce je rozdělena do dvou částí. V úvodním literárním přehledu jsou definovány základní termíny z odborné literatury, které řeší základní pojmy z teorie svazů, uzávěrové operátory a větu o pevném bodě.

První kapitola v praktické části diplomové práce se zabývá formální konceptuální analýzou a reprezentační větou pomocí Galoisových konexí. Jsou zde vysvětleny pojmy jako formální kontext, formální koncept nebo konceptuální svaz, které jsou aplikovány na konkrétních příkladech v poslední kapitole. Další kapitola pak představuje jednotlivé internetové vyhledávače a principy jejich činnosti. Vyhledávače lze obecně rozdělit na tři základní typy. Jsou to vyhledávače katalogové, fulltextové a meta vyhledávače.

Jako první se na Internetu objevily katalogové vyhledávače, které do jisté míry fungují dodnes. S rozšiřováním Internetu vznikaly nové, modernější způsoby vyhledávání, mezi které patří fulltextové vyhledávání a meta vyhledávání. Katalogové vyhledávače obsahují velkou databázi odkazů na různé webové stránky celého Internetu, uspořádanou do stromové struktury a řazenou do jednotlivých kategorií a podkategorií podle tematických sekcí. Fulltextové vyhledávače automaticky prochází webové stránky na základě určených algoritmů. Kvalitu procházených a stahovaných stránek vyhodnotí na základě mnoha faktorů. Meta vyhledávače nemají vlastní databázi webových stránek, ale používají a spoléhají se při zadání dotazu na jiné vyhledávače.

Poslední kapitola aplikuje metodu formální konceptuální analýzy na konkrétních příkladech. První příklad se zabývá srovnáním šesti českých internetových vyhledávačů podle předem zvolených kritérií, kterými byli unikátní návštěvníci, typ vyhledávače, vyhledání dokumentů, více jazyků, třídění výsledků, náhled webové stránky a vyhledání videa a obrázků. V druhém příkladu zůstaly zachovány stejné atributy, pouze dochází ke srovnání sedmi zahraničních internetových vyhledávačů. Poslední příklad se pak zaměřuje na porovnání českých internetových vyhledávačů právě s těmi zahraničními. Pro objektivnost srovnávání zůstaly atributy z předchozích příkladů nezměněné. Z výsledků je patrné, že při hledání na Internetu volí většina uživatelů fulltextové vyhledávače, protože při správném zadání dotazu mají relativně vysokou šanci, že najdou požadovanou odpověď na zadaný dotaz, tedy konkrétní webovou stránku. Katalogové vyhledávače spravuje webmaster, u kterého se musejí uživatele spolehnout na pečlivé zadání webové stránky, kterou přidal do katalogu. U meta vyhledávačů si uživatelé nemohou být nikdy jistí, od jakého vyhledávače výsledný dotaz vzešel, popřípadě jak kvalitní nalezená odpověď je. Proto dnes většina portálů přistoupila na kombinovanou formu vyhledávání, což znamená vyhledávání jak v katalogu, tak i nějakou fulltextovou technologií.

V literatuře je popsána celá řada příkladů použití formální konceptuální analýzy na reálných datech. Tato metoda se běžně používá v informatice, knihovnictví, lékařství, psychologii, stavebnictví či biologii.

ZÁVĚR V ANGLIČTINĚ

In information science, formal concept analysis is the method of deriving concepts within a certain hierarchy. Each concept in this hierarchy represents a set of objects that share the same values for a particular set of characteristics, and each sub-concept in the hierarchy contains a subset of the other objects in the same conceptions as its superior set of objects. The original purpose of formal concept analysis was concrete representation of complete lattice and their properties using data tables that present binary relations between objects and attributes. This analysis was published by Rudolf Wille in 1984 and continued the work of Birkhoff and others in 1930.

The aim of this thesis was to compile the basics of formal conceptual analysis and formulate the basic representation theorem using Galois´s connections, provide an overview of the various search engines, to describe the basic principles of their activities and to analyze search engines and provided services just using methods of formal conceptual analysis.

The work is divided into two parts. In the introductory literature review the basic terms from technical literature are defined that deal with the basic concepts of lattice theory, closure operators and fixed point theorem.

The first chapter in the practical part of the thesis deals with the formal conceptual analysis and with representation sentence using Galois´s connections. Concepts as a formal context, formal concept or conceptual lattice are explained here and these concepts are applied to specific examples in the last chapter. The next chapter then describes individual search engines and the principles of their activities. Search engines can be in general divided into three basic types. These are the catalog search engines, full-text and meta search engines.

As the first, catalog search engines appeared on the Internet which have been working till today to some extent. With the expansion of the Internet new, more modern methods of searching was developed, which included fulltext search and meta search. Catalog search engines contain a large database of links to various websites around the Internet, which is arranged in a tree structure and sorted into individual categories and subcategories according to thematic sections. Full-text search engines automatically browse the Web sites according to specific algorithms. The quality of browsed and downloaded pages is evaluated based on many factors. Meta search engines have their own database of web pages, but by querying use and rely to other search engines.

The last chapter applies the method of formal concept analysis to concrete examples. The first example deals with a comparison of six Czech search engines according to preselected criteria, which were unique visitors, browser type, locate documents, multiple languages, classification of results, preview site and find videos and pictures. In the second example, the same attributes retained, only seven foreign search engines are compared, according to these attributes. The last example is then focused on a comparison of Czech Internet search engines with just those foreign. Attributes remain unchanged from previous examples for comparison objectivity. The results show that when searching the Internet, most users choose full-text search, because when the question is specified correctly, the user has a relatively high chance of finding the answer to his query, thus the specific Web page. Catalog Search Engine manages the webmaster, so that the user must rely on carefully added address of web page that the webmaster added to the catalog. When using the meta search engine, users can never be sure what search engine the query comes from or how good the answer is. Therefore, today most of the portals use a combined search methods, which means to search by the catalog, as well as a full-text technologies.

The literature describes many examples of the use of formal concept analysis with real data. This method is commonly used in information science, library science, medicine, psychology, civil engineering or biology.

SEZNAM POUŽITÉ LITERATURY

Knižní publikace

[1] WILLE, R., GANTER, B. Formal Concept Analysis – Mathematical Foundations. 1st ed. Springer, 1998. 284 s. ISBN 3-540-62771-5.

[2] CHAJDA, Ivan. Algebra 3. Olomouc: Univerzita Palackého, 1998. 125 s. ISBN 80-7067-803-8.

[3] KOPKA, Jan. Svazy a booleovy algebry. Ústí nad Labem: Univerzita J. E. Purkyně, 1991. 244 s. ISBN 80-7044-025-2.

[4] RACHŮNEK, Jiří. Svazy. Olomouc: Univerzita Palackého, 2003. 85 s. ISBN 80-2440-650-0.

Internetové zdroje

[5] HÝSEK, Jiří. *Semimodulární, modulární a distributivní svazy* [online]. 2008-01-17 [cit. 2012-03-02]. Dostupný z: <http://trace.dump.cz/papers/qm3.pdf>.

[6] KUČERA, Radan. *Základy teorie svazů* [online]. 2010 [cit. 2012-03-05]. Dostupné z: <http://www.math.muni.cz/~kucera/texty/Svazy2010.pdf>.

[7] Homomorfismus. *Wikipedia* [online]. 2012 [cit. 2012-04-05]. Dostupné z: <http://cs.wikipedia.org/wiki/Homomorfismus>.

[8] BĚLOHLÁVEK, Radim. *Konceptuální svazy a formální konceptuální analýza* [online]. 2004 [cit. 2012-03-17]. Dostupné z: <http://belohlavek.inf.upol.cz/publications/Bel_Ksfka.pdf>.

[9] Vyhledávače, jejich typy a rozdělení. *SEO master* [online]. 2012 [cit. 2012-04-05]. Dostupný z: <http://www.seomaster.cz/seznam-vyhledavacu-a-popis>.

[10] Roboti a crawleři aneb jak fungují fulltextové vyhledávače. *Ataxo* [online]. 2010 [cit. 2012-04-05]. Dostupné z: <http://www.ataxo.cz/informace/vyhledavace-katalogy/vyhledavace>.

[11] Vyhledávače. *Jak psát web* [online]. 2012 [cit. 2012-04-27]. Dostupný z: <http://www.jakpsatweb.cz/vyhledavace.html>.

[12] Jak vyhledávač pracuje. *Internetové vyhledávání* [online]. 2012 [cit. 2012-04-27]. Dostupný z: <http://www.vyhledej.estranky.cz/clanky/jak-vyhledavac-pracuje.html>.

[13] Seznam – nejpoužívanější český vyhledávač. *Ataxo* [online]. 2012 [cit. 2012-05-02]. Dostupný z: <http://www.ataxo.cz/informace/vyhledavace-katalogy/seznam>.

[14] Seznam. *Wikipedia* [online]. 2012 [cit. 2012-04-05]. Dostupné z: <http://cs.wikipedia.org/wiki/Seznam.cz>.

[15] Centrum. *Wikipedia* [online]. 2012 [cit. 2012-04-05]. Dostupné z: <http://cs.wikipedia.org/wiki/Centrum.cz>.

[16] Jyxo. *Wikipedia* [online]. 2012 [cit. 2012-04-05]. Dostupné z: http://cs.wikipedia.org/wiki/Jyxo.cz

[17] Informace. *Jyxo* [online]. 2012 [cit. 2012-05-07]. Dostupné z: <http://jyxo.cz/d/info>.

[18] Google. *Wikipedia* [online]. 2012 [cit. 2012-04-05]. Dostupné z: <http://cs.wikipedia.org/wiki/Google>.

[19] Přehled technologie. *Google* [online]. 2008 [cit. 2012-05-12]. Dostupné z: <http://www.google.cz/intl/cs/about/corporate/company/tech.html>.

[20] Bing. *Wikipedia* [online]. 2012 [cit. 2012-04-05]. Dostupné z: <http://cs.wikipedia.org/wiki/Bing>.

[21] Google – světová jednička. *Ataxo* [online]. 2010 [cit. 2012-05-12]. Dostupné z: <http://www.ataxo.cz/informace/vyhledavace-katalogy/google>.

[22] Katalogové vyhledávače. *Srovnání vyhledávacích serverů* [online]. 2012 [cit. 2012-05-14]. Dostupné z: <http://vyhledavace.zkrat.net/katalog.html>.

[23] Katalogy. *Jak psát web* [online]. 2012 [cit. 2012-04-27]. Dostupný z: <http://www.jakpsatweb.cz/katalogy.html>.

[24] Yahoo. *Ataxo* [online]. 2010 [cit. 2012-05-05]. Dostupné z: <http://www.ataxo.cz/informace/vyhledavace-katalogy/yahoo>.

[25] Historie Yahoo!. *CIO Business world* [online]. 2010 [cit. 2012-05-05]. Dostupné z: <http://businessworld.cz/ostatni/historie-yahoo-5842>.

[26] Open Directory Project (ODP) – DMOZ. *DMOZ* [online]. 2010 [cit. 2012-05-09]. Dostupné z: <http://www.dmoz.cz/>.

[27] META vyhledávače. *Srovnání vyhledávacích serverů* [online]. 2012 [cit. 2012-05-06]. Dostupné z: <http://vyhledavace.zkrat.net/meta.html>.

[28] České vyhledávače. *Národní technická knihovna* [online]. 2012 [cit. 2012-04-26]. Dostupné z: <http://www.techlib.cz/cs/227-ceske-vyhledavace/>.

[29] Vyhledávací nástroje internetu. *Infogram* [online]. 2012 [cit. 2012-04-26]. Dostupné z: <http://www.infogram.cz/article.do?articleId=1761>.

[30] *Siteinfo.cz* [online]. 2012 [cit. 2012-05-06]. Dostupné z: <http://www.siteinfo.cz/>.

[31] *Compete* [online]. 2012 [cit. 2012-05-06]. Dostupné z: <http://www.compete.com/us/>.

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

FCA	Formal Concept Analysis.
PR	Page Rank
HTML	Hyper Text Markup Language.
WWW	World Wide Web.

SEZNAM OBRÁZKŮ

SEZNAM TABULEK