

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A
KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV AUTOMATIZACE A MĚŘÍCÍ TECHNIKY**



**FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF CONTROL AND INSTRUMENTATION**

METODY REPREZENTACE ZNALOSTÍ

KNOWLEDGE REPRESENTATION METHODS

BAKALÁŘSKÁ PRÁCE
BACHELORS'S THESIS

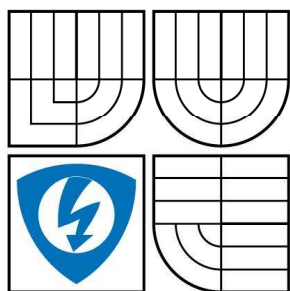
AUTOR PRÁCE
AUTHOR

JOSEF VERBÍK

VEDOUČÍ PRÁCE
SUPERVISOR

doc. Ing. VÁCLAV JIRSÍK, CSc.

BRNO 2008



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav automatizace a měřicí
techniky

Bakalářská práce

bakalářský studijní obor

Automatizační a měřicí technika

Student: Verbík Josef

ID: 83155

Ročník: 3

Akademický rok: 2007/2008

NÁZEV TÉMATU:

Metody reprezentace znalostí

POKYNY PRO VYPRACOVÁNÍ:

1. Na základě rešerše popište jednotlivé reprezentace znalostí pro expertní systémy.
2. Vhodně si zvolte kritéria a kvalifikační faktory pro srovnání jednotlivých reprezentací znalostí.
3. U každé reprezentace znalostí uveďte a zdůvodněte příklad pro nejvhodnější úlohu.

DOPORUČENÁ LITERATURA:

- [1] Berka, P. a kol.: Expertní systémy. VŠE Praha, 1998
- [2] Dvořák, J.: Expertní systémy. VUT Brno, 2004
- [3] Mařík V., Štěpánková O., Lažanský J.: Umělá inteligence (1). Academia, Praha 1993
- [4] Mařík V., Štěpánková O., Lažanský J.: Umělá inteligence (2). Academia, Praha 1997

Termín zadání: 1.2.2008

Termín odevzdání: 6.6.2008

Vedoucí práce: doc. Ing. Václav Jirsík, CSc.

prof. Ing. Pavel Jura, CSc.

předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Abstrakt

Účelem této práce je rozbor metod reprezentace znalostí v expertních systémech. Práce popisuje druhy expertních systémů a jejich části. Dále jsou popsány požadavky pro efektivní reprezentaci znalostí. Práce se zabývá čtveřicí metod reprezentace znalostí. Jsou to pravidla, rámce, sémantické sítě a predikátová logika. U každé metody jsou popsány její charakteristické vlastnosti a obecné výhody či nevýhody.

Pro samotné zhodnocení metod byl použit expertní systém NPS32. Ke každé z metod byla vytvořena báze znalostí charakterizující jejich klady a zápory.

Klíčová slova: expertní systém, reprezentace znalostí, báze znalostí, sémantická síť, rámce, pravidla, predikátová logika

Abstract

The point of this thesis is to analyse methods of knowledge representation within expert systems. This thesis describes the various kinds of expert systems and their components. It also describes the requirements necessary for efficient knowledge representation. The thesis deals with the following four methods of knowledge representation: rules, frames, semantic nets and predicate logic. For each of these methods, a description of its characteristic features is provided together with a list of general advantages and disadvantages.

For the actual evaluation of the methods, the NPS32 expert system has been used. For each method, a knowledge platform has been created featuring their pros and cons.

Key words: expert system, knowledge representation, knowledge base, semantic net, frames, rules, predicate logic

Bibliografická citace

VERBÍK, J. *Metody reprezentace znalostí*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 54 s. Vedoucí bakalářské práce doc. Ing. Václav Jirsík, CSc.

P r o h l á š e n í

„Prohlašuji, že svou diplomovou práci na téma "Metody reprezentace znalostí" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.“

V Brně dne : 2.6.2008

Podpis:

P o d ě k o v á n í

Děkuji vedoucímu bakalářské práce doc. Ing. Václavu Jirsíkovi, CSc. za inspiraci a cenné rady při vypracování bakalářské práce.

V Brně dne : 2.6.2008

Podpis:

OBSAH

OBSAH.....	7
1. ÚVOD	9
2. EXPERTNÍ SYSTÉMY	10
2.1 Struktura.....	12
2.1.1 Báze znalostí.....	13
2.1.2 Inferenční mechanismus	14
2.2 Druhy ES.....	16
2.2.1 Diagnostické ES.....	16
2.2.2 Plánovací ES	17
2.2.3 Hybridní ES	18
2.2.4 Prázdné ES.....	19
2.3 Historie ES	19
3. REPREZENTACE ZNALOSTÍ	22
3.1 Význam znalostí.....	22
3.1.1 Logika	24
3.1.2 Deduktivní logika	25
3.1.3 Výroková logika	26
3.2 Predikátová logika	27
3.3 Sémantické sítě	29
3.4 Rámce	30
3.5 Pravidla	32
4. EXPERTNÍ SYSTÉM NPS32	34
4.1 Matematický aparát.....	34
4.2 Báze znalostí	35
5. SROVNÁNÍ RZ	37
5.1 Rámce	37
5.1.1 Zhodnocení	40
5.2 Sémantické sítě	40
5.2.1 Zhodnocení	44

5.3 Pravidla	44
5.3.1 Zhodnocení	46
5.4 Predikátová logika	46
5.4.1 Zhodnocení	46
6. ZÁVĚR	48
7. SEZNAM OBRÁZKŮ A TABULEK	49
8. POUŽITÁ LITERATURA	50

1. ÚVOD

Dnešní doba je označována za informační. Informace hýbou světem, kdo k nim má přístup a umí je použít, má peníze a moc. Klade se tedy důraz na získávání a zpracovávání informací. K tomuto účelu mimo jiné slouží i expertní systémy. Informace uložené v bázi znalostí jsou reprezentovány určitými metodami.

Cílem této bakalářské práce je srovnání základních metod reprezentace znalostí v expertních systémech. Úvodní část popisuje expertní systém a jeho části, zejména bázi znalostí a inferenční mechanismus. Dále jsou popsány jednotlivé druhy expertních systémů a jejich historie.

V další kapitole jsou uvedeny požadavky pro efektivní reprezentaci znalostí v expertních systémech. Dále jsou charakterizovány jednotlivé metody reprezentací znalostí a jejich výhody či nevýhody.

K samotnému porovnání a zhodnocení metod reprezentace znalostí v expertních systémech posloužil expertní systém NPS32. Na něm jsou předvedeny názorné ukázky výhod či nevýhod jednotlivých metod.

2. EXPERTNÍ SYSTÉMY

Expertní systémy (ES) jsou softwarové nástroje, které mají za úkol poskytovat expertní rady, rozhodnutí nebo doporučit řešení v konkrétní situaci [Dvo04]. Jsou navrženy tak, že mohou zpracovávat neurčité a nenumerní informace. Expertní systémy nahrazují algoritmy, které nejsou schopné vyřešit určité problémy tradičními postupy.

ES nahrazují experta (nejlépe více expertů) v dané oblasti a pomáhají řešit dané problémy.

Charakteristické rysy expertních systémů jsou [Bkol98]:

- oddělení báze znalostí od inferenčního mechanismu - tímto se liší expertní systémy od klasických programů,
- schopnost rozhodování při neurčitosti,
- schopnost vysvětlování,
- dialogový režim.

Vysvětlovací činnost zvyšuje důvěru v závěry a doporučení expertního systému. Obvykle vysvětluje právě položený dotaz, znalosti relevantní k nějakému tvrzení, právě zkoumanou cílovou hypotézu, právě probíhající odvozování [Hab04].

Dialogový režim usnadňuje komunikaci laika s expertním systémem. Uživatel komunikuje se systémem formou „otázka systému – odpověď uživatele“. Dialog uživatele se systémem má podobu dialogu laika nebo méně zkušeného odborníka s expertem. Systém se ptá na údaje, které souvisí s konzultovaným problémem, a na základě odpovědí dochází k závěru či doporučení [Hab04].

Expertní systémy vynikají nad lidskými experty zejména v [PK99]:

- Dostupnosti hardware – dnešní doba je prošpikována počítači a jinou výpočetní technikou jako jsou kapesní počítače (PDA) a „inteligentní“ mobilní telefony (nahrazují PDA), na kterých lze ES provozovat.
- Časové dostupnosti – expertízu lze získat kdykoli.

- Ceně – cena expertízy je pro uživatele (zákazníka) nižší než s lidským expertem.
- Rychlosti – ES může rychleji reagovat na vstupní požadavky (v závislosti na použitém software a hardware).
- Spolehlivosti – ES vytvořený více experty eliminuje případné chyby či konfliktní situace.
- Vícenásobné expertíze – při řešení problému lze využít více výstupů expertních systémů zaměřených na různé oblasti.
- Vysvětlení a trénování – ES dokáže přesně vysvětlit postup, jakým přišel k výsledku svého řešení.
- Stálosti – časově neměnná a úplná odpověď na stejný problém.

Expertní systémy mají samozřejmě i nevýhody [ŠtýW]. První nevýhodou je nebezpečí selhání při změněných podmínkách. Tj. když například při výběru vhodného produktu se změní podstatná vlastnost produktu (cena). Proto není vhodné používat ES na výběr produktů s rychle měnící se cenou (například výpočetní technika). Další nevýhodou ES je neschopnost rozpoznat své hranice použitelnosti. ES taky nedokáže nahradit „šestý smysl“ (intuice) experta získaný praxí a zkušenostmi.

ES mají široký rozsah použití [PK99]. Typické oblasti použití jsou:

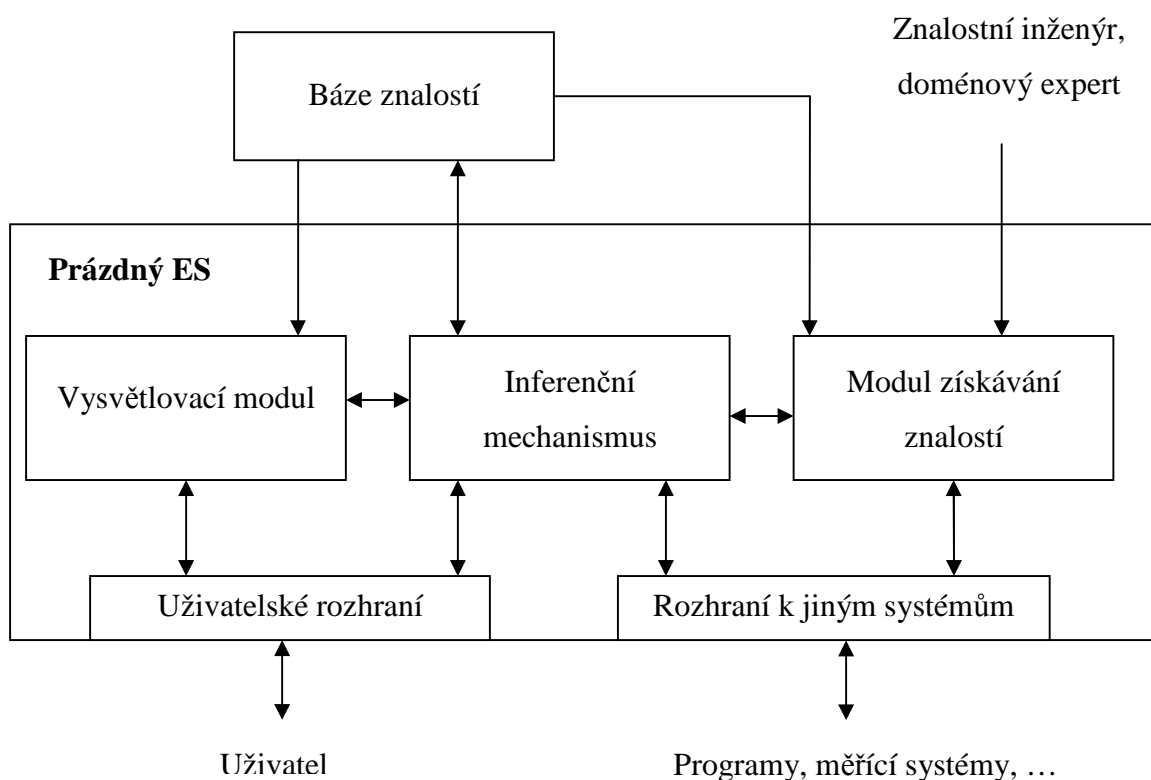
- konfigurace – výběr a sestavení vhodných komponent systému,
- diagnostika – odvození závěru na základě dodaných faktů (např. v lékařství),
- monitorování – porovnání měřených dat s daty očekávanými,
- ladění – zhotovení pravidel pro odstranění poruch systému,
- interpretace – vysvětlení získaných dat,
- plánování – navržení posloupností činností k řešení daného problému,
- prognózování – předpověď výsledku dané situace,
- řízení – regulace procesů (může obsahovat všechny předchozí typy),

- učení – inteligentní výuka, při které se uživatel smí ptát (co?, jak?, proč?, ...).

2.1 STRUKTURA

Expertní systém [Dvo04] se skládá z těchto hlavních částí (Obr. 1):

- báze znalostí,
- inferenční (řídící) mechanismus,
- I/O rozhraní (uživatelské, vazby na jiné systémy),
- vysvětlovací modul,
- modul pro získávání znalostí.



Obr. 1: Architektura expertního systému

2.1.1 Báze znalostí

V bázi znalostí jsou soustředěny veškeré znalosti experta, které jsou potřebné k řešení daného problému. Je to jediné místo, kde jsou znalosti ukládány, podléhají určité organizaci. Báze by měla být modulární (tj. aby bylo možné poznatky v bázi znalostí kdykoliv doplnit anebo aktualizovat tak, aby báze stále odpovídala úrovni nejnovějších poznatků v dané oblasti) a otevřená pro zahrnutí nových znalostí a případnou eliminaci neefektivních či redundantních znalostí. Znalosti jsou v podstatě pravidla, kterými se systém řídí.

V bázi znalostí je zapsáno velké množství různých znalostí:

- od nejobecnějších až po úzce odborné,
- od učebnicových až po soukromé znalosti,
- od exaktně (založený na přesnosti, přesný, dokonalý) prokázaných znalostí až k nejistým heuristikám (teorie řešení problémů),
- od jednoduchých znalostí až po metaznalosti (znalosti o znalostech).

K reprezentaci znalostí se používá řada technik:

- matematická logika,
- konceptuální grafy,
- rozhodovací stromy,
- pravidla,
- rámce a scénáře,
- sémantické sítě,
- objekty,
- atd.

Znalosti se mohou dělit na mělké a hluboké. Mělké znalosti jsou založeny na heuristických a empiristických (zkušenostních) znalostech. Hluboké znalosti jsou založeny na strukturách, funkcích a vlastnostech objektů.

Během řešení problému se vytváří báze faktů, která obsahuje data (vstupní data a průběžné výsledky) související s řešeným problémem.

2.1.2 Inferenční mechanismus

Inferenční mechanismus je programový modul, který předem udává strategii využívání znalostí z báze znalostí, zprostředkovává komunikaci mezi bází znalostí a bází dat (respektive bází znalostí a uživatelem expertního systému).

Typický inferenční mechanismus je založen na:

- inferenčním pravidle pro odvozování nových poznatků z existujících znalostí,
- strategii prohledávání báze znalostí.

Přehled metod inference (usuzování, odvozování určitých výroků z jiných):

- dedukce – logické usuzování (závěry musí vyplývat z předpokladů),
- indukce – postup od specifického případu k obecnému,
- abdukce – usuzování směřující ze správného závěru k předpokladům, které jej mohli způsobit,
- heuristiky – pravidla založená na zkušenostech,
- generování a testování – metoda „pokus – omyl“,
- analogie – odvození závěru podle podobné situace,
- defaultní inference – usuzování na základě obecných znalostí v případě absence specifických znalostí,
- nemonotónní inference – oprava dosavadních znalostí na základě nové informace,
- intuice – téměř nevysvětlitelný způsob usuzování, které se opírá o nevědomé rozpoznání nějakého vzoru. Tento typ usuzování nebyl v expertních systémech ještě aplikován. Je možné, že se v budoucnu bude využívat v usuzování neuronových sítí.

2.1.2.1 Neurčitosti

Důležitou schopností inferenčního mechanismu je zpracování neurčitosti [ŠtýW]. Na neurčitost se z pohledu znalostního systému díváme ze dvou hledisek:

- Jako na neurčitost v bázi dat – vstupní data, která na začátku, respektive v průběhu odvozování, vkládá inferenční mechanismus do báze faktů, i ty, které jsou odvozené, mohou být neurčité.
- Jako na neurčitost v bázi znalostí – ne každý závěr je možné z předpokladu odvodit se stoprocentní platností.

Neurčitost v systému má mnohé příčiny. Ať už je to nepřesnost měřících přístrojů, nebo vágní jazyková formulace, nebo subjektivní dojem, se všemi je potřeba pracovat. Právě zpracování neurčitosti je jednou z nejpodstatnějších složek expertních systémů.

Nejistota, neurčitost bývá v expertních systémech vyjadřována různými způsoby:

- mohou to být různé váhy, míry, stupně důvěry aj. nazývané a formulované subjektivní pravděpodobností. Tyto numerické parametry jsou přiřazeny jednotlivým tvrzením, pravidlům a datům. Nástrojem v těchto podmínkách je matematická pravděpodobnost.
- je-li k reprezentaci využito přirozeného jazyka, pak je neurčitost dána stupněm vágnosti použitých jazykových termínů (velmi vysoký, asi ano atd.). K formalizaci vágních pojmů se využívá fuzzy množinové matematiky, inferenční mechanismy pak využívají principů fuzzy jazykové logiky.

Různé typy chyb způsobující neurčitost:

- dvojznačnost – informace je interpretována dvěma či více způsoby,
- náhodná chyba – kolísání v měření kolem určité střední hodnoty,
- systematická chyba – zkresluje výsledek měření určitým způsobem a s jistou pravidelností,
- nekompletnost – nedostatek informací,
- nekorektnost – významově špatná informace
 - lidská chyba – špatné odečtení hodnoty,
 - strojová chyba – nespolehlivost zařízení,

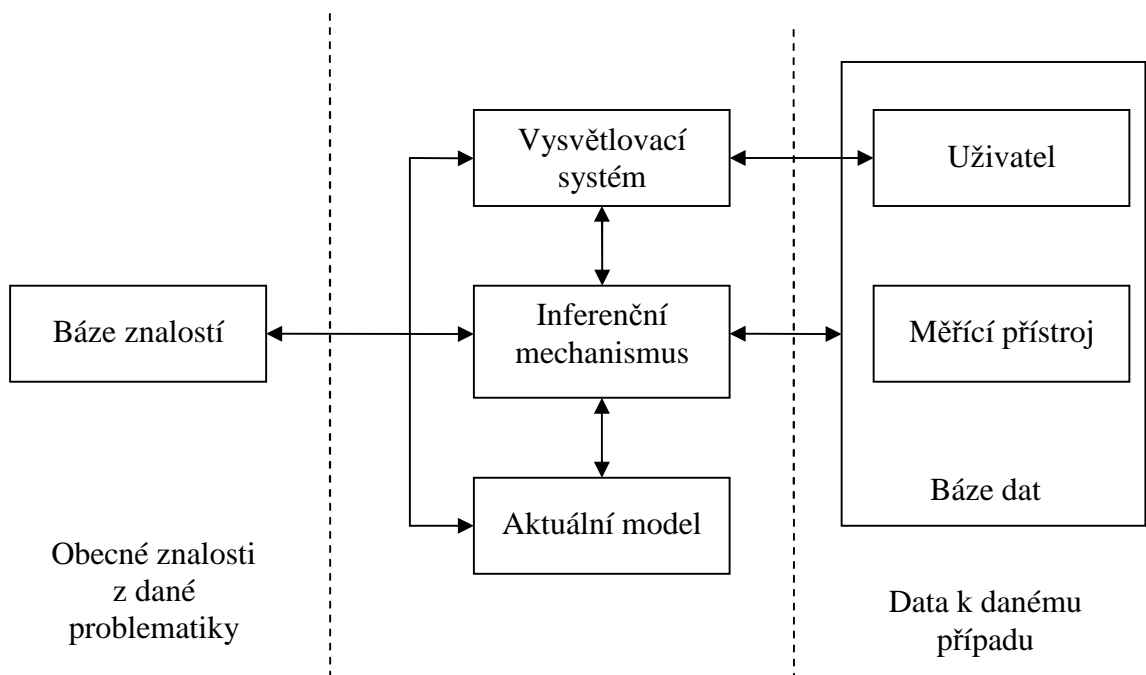
- chyba měření,
- chyba usuzování
 - induktivní chyba (výrok není možno stoprocentně dokázat),
 - deduktivní chyba (stejný výrok může být platný i při negaci některé z premis).

2.2 DRUHY ES

ES systémy se dělí podle charakteru řešených úloh na diagnostické (tvoří až 90% všech ES), plánovací, hybridní a prázdné [RadW].

2.2.1 Diagnostické ES

Jsou určeny pro efektivní interpretaci dat s cílem určit, která z hypotéz (jsou předem dány) se nejlépe shoduje s aktuálními daty [MŠL97]. Nejtypičtější příklad je stanovení diagnózy pacienta na základě jeho subjektivních potíží (jsou dána možná řešení a jedno z nich systém vybere). Typická architektura diagnostických expertních systémů je znázorněna na Obr. 2.



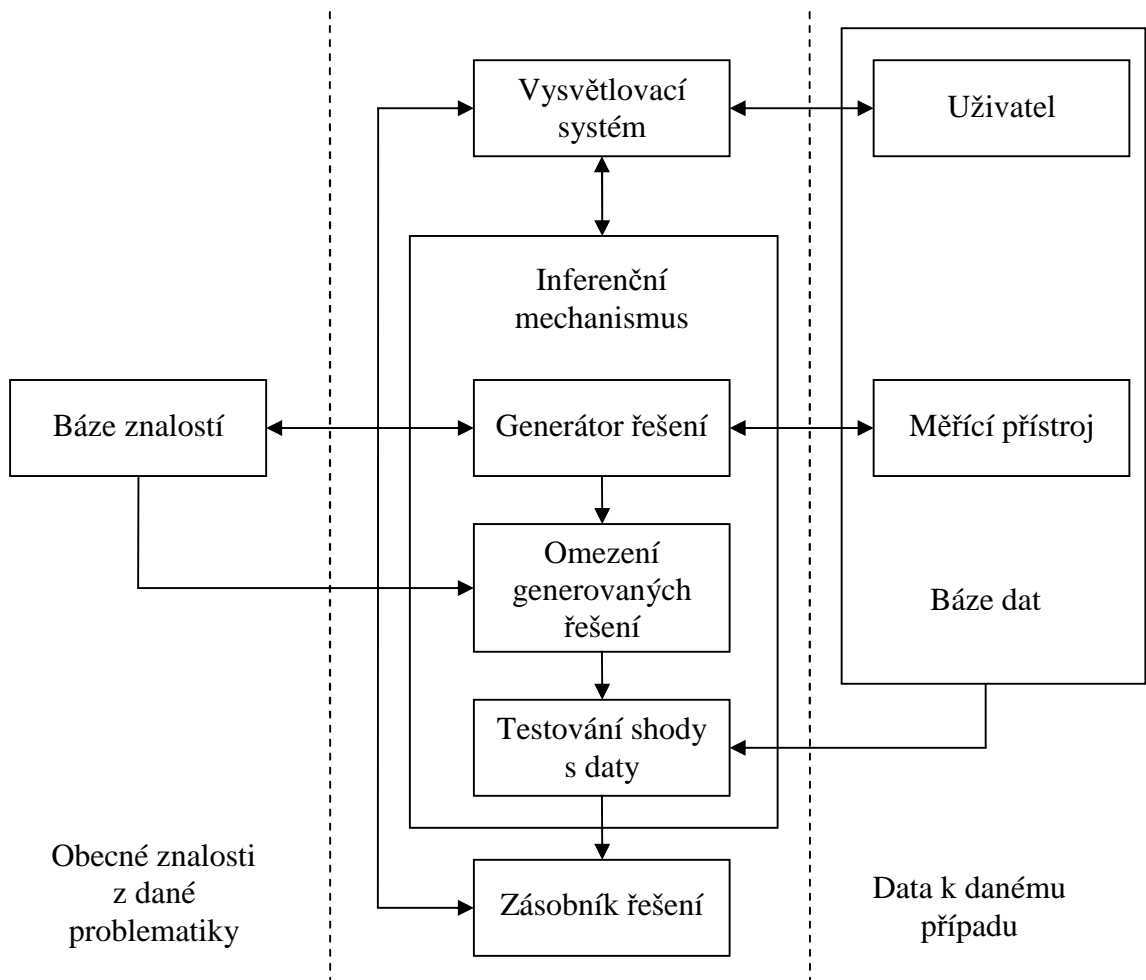
Obr. 2: Architektura diagnostických ES

Báze dat může být složena jak z uživatelem zadaných hodnot, tak z naměřených hodnot (subjektivní pocit pacienta a jeho naměřená teplota). Na počátku jsou data vložena expertem, poté jsou tato data v průběhu odvozování měněna.

Vysvětlovací systém uchovává a vysvětluje postup, jakým bylo dosaženo výsledku.

2.2.2 Plánovací ES

Jsou určeny pro řešení úloh, při kterých je znám stav počátečního objektu a cíl řešení [MŠL97]. Systém má s využitím dat a zadáním nalézt posloupnost povolených kroků. Důležité je také nalezení (pokud je to možné) optimální posloupnosti. Typická architektura plánovacích expertních systémů je znázorněna na Obr. 3.



Obr. 3: Architektura plánovacích ES

Hlavní částí plánovacích expertních systémů je generátor možných řešení. Generátor je omezen pravidly z báze znalostí. Dále je testována shoda řešení s daty z báze dat. Výsledkem je seznam možných řešení, která jsou ohodnocena stupněm kvality.

2.2.3 Hybridní ES

Hybridní expertní systémy mají kombinovanou architekturu, tj. mohou kombinovat plánovací ES, diagnostické ES, neuronové sítě či evoluční metody. K hybridním systémům patří například výukový systém, který diagnostikuje znalosti studenta a podle toho plánuje jeho další výuku.

2.2.4 Prázdné ES

Jejich podstata vychází ze skutečnosti, že základní a zcela univerzální částí expertního systému je jeho inferenční mechanismus. Ten může pracovat s bázemi znalostí, které mají společnou architekturu, ale mohou být problémově orientované na jiné oblasti problému. Prázdný ES obsahuje inferenční mechanismus a další nezbytné součásti, pouze báze znalostí je prázdná. Doplněním báze znalostí k prázdnému systému se systém orientuje na danou problematiku. Dodáním báze dat je pak vždy řešen konkrétní případ.

Prázdné ES jsou relativně levné, protože nejdražší je vytvoření vyladěné báze znalostí expertem na danou problematiku.

2.3 HISTORIE ES

Mezi lety 1965 a 1970 byly expertní systémy v počáteční fázi vývoje a výzkumu [MŠL93]. V těchto letech vznikly expertní systémy DENDRAL a MACSYMA. DENDRAL je typický plánovací systém, který je určen k odvození struktur chemických látek. DENDRAL je nejdéle využívaný ES v praxi. Byl vytvořen dokonce dříve než se zavedl pojem expertní systém. Systém MACSYMA je stále využíván při řešení rovnic o velkém počtu, protože obsahuje soubor nástrojů pro manipulaci s matematickými výrazy a vzorci.

V letech 1970 až 1975 (etapa výzkumných prototypů) vznikly první dva úspěšné diagnostické expertní systémy MYCIN a PROSPECTOR. Oba dva umí pracovat s neurčitostí při usuzování. MYCIN diagnostikoval v medicíně infekční onemocnění krve. PROSPECTOR sloužil k odhalení rudných ložisek z dostupných geologických dat. Proslavil se tím, že během prvních pár týdnů svého provozu odhalil velké ložisko rud v hodnotě sta milionů dolarů.

Mezi lety 1975 až 1981 (etapa experimentálního nasazování) vznikly systémy PUFF, SACON, INTERNIST a mnoho dalších. Systém PUFF sloužil ke konzultacím týkajících se možných příčin instruktivních potíží dýchacích cest. Byl realizován pomocí EMYCIN (prázdný systém MYCIN). Systém SACON sloužil uživatelům softwarového balíku MARC. SACON využívá architekturu MYCIN. Systém INTERNIST (později přejmenován na CADUCEUS) byl vyvíjen pro celou oblast

interního lékařství. Je považován za jeden z nejrozsáhlejších systémů v historii. Systém se dodnes v medicíně využívá.

V roce 1981 byl poprvé použit ke komerčnímu využívání expertní systém XCON. Sloužil ke konfiguraci počítačů v jisté firmě. Úspora prý dosahovala až k deseti milionům dolarů ročně. Dalším úspěšným systémem je DIPMETER ADVISOR [Hab04]. Sloužil pro potřeby naftařských společností (rady pro obsluhu vrtných zařízení, rady při problémech vrtů). Náklady na vývoj tohoto systému byly vyčísleny na 21,5 milionu dolarů. Po nasazení tohoto systému se však firmám po několika měsících vrátily.

Oblast využití	Příklady systémů
Medicína	MYCIN, PIP, PUFF, MEDICO, PROTIS, HEADMED, NEUREX, EEG, CASNET (GLAUCOMA), ONCOCIN
Chemie	DENDRAL, SECS, SYNCHEM
Genetika	MOLGEN
Geologie	PROSPECTOR
Mechanika	SACON, MECHANO
Matematika	AM, MASCYMA
Pedagogika	WHY, SOPHIE, GUIDON, BLAH
Právo	LEGOL, TAXMAN, LRS, MATRIM

Tab. 1: Přehled expertních systémů dle oblastí využití

V dnešní době se ES vyvíjí dvěma směry [Dvo04]. Mluví se o první a druhé generaci. Charakteristické rysy první generace jsou:

- jeden způsob reprezentace znalostí,
- malé schopnosti vysvětlování,
- znalosti pouze od expertů.

Druhá generace se vyznačuje:

- modulární a víceúrovňovou bází znalostí,
- hybridní reprezentací znalostí,
- zlepšeným vysvětlovacím mechanismem
- prostředky pro automatizované získávání znalostí.

ES první generace se vylepšují přidáváním různých modulů [MŠL93]. Jsou to například moduly pro přímou komunikaci s externími programy a moduly pro propojení s databázovými systémy.

3. REPREZENTACE ZNALOSTÍ

V reprezentaci znalostí je nejdůležitější vlastností efektivní reprezentace [MŠL97]. Požadavky na reprezentaci znalostí jsou:

- dostatečně přirozená reprezentace pro danou oblast a její expresivita (různým způsobem vyjádřené),
- umožnění aplikace efektivních deduktivních prostředků,
- rychlý přístup k bázi znalostí a bázi dat.

Dalším požadavkem je také modularita báze znalostí [RadW]. Do již existující báze je třeba doplňovat nové znalosti, odstranit nepotřebné nebo upravovat stávající. Modularita přináší výhody v organizaci. Nevýhodou je větší složitost prohledávání, protože údaje objektu nemusí být soustředěny na jednom místě (nemusí zohledňovat sémantiku báze). Sémantické sdružování znalostí pak vyplývá především z požadavku na rychlost vybavování znalostí a z potřeby vytvářet hierarchii pojmů. V konkrétních případech se musí zvolit kompromis mezi oběma požadavky.

3.1 VÝZNAM ZNALOSTÍ

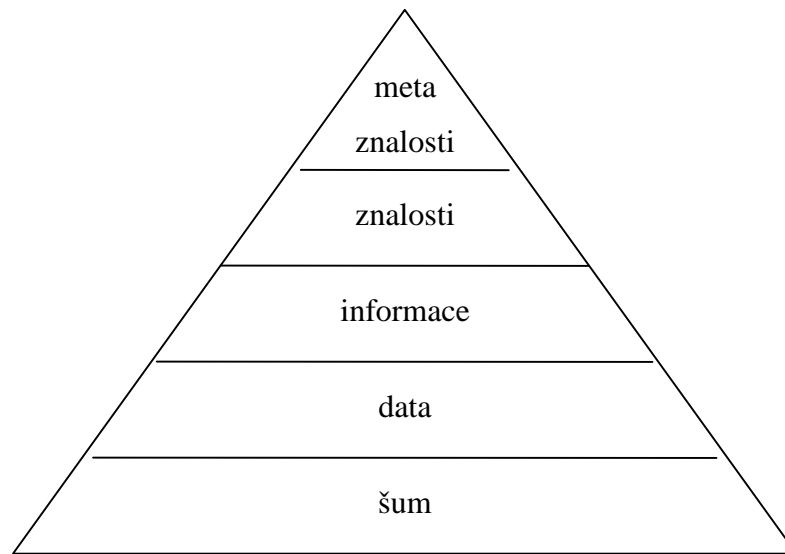
Reprezentace znalostí se v systémech umělé inteligence dělí na [MŠL97]:

- deklarativní – vyjadřují, co je nebo má být poznáno či dokázáno,
- procedurální – vyjadřují, jak poznávat nebo odvozovat.

Deklarativně reprezentované znalosti se nazývají poznatky, procedurálně reprezentované mají podobu pravidel. Stejná znalost může být reprezentována jako poznatek („myš je savec“) nebo jako pravidlo („Je-li X myš, potom je X savec“).

Reálné systémy v sobě obsahují oba typy reprezentace. Většinou je velmi obtížné přesně stanovit hranici a jasně oddělit procedurálně a deklarativně reprezentované znalosti.

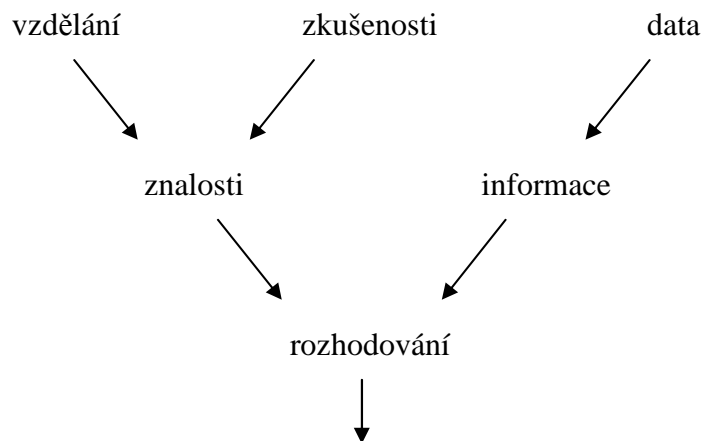
Znalosti v expertních systémech mají nejvyšší prioritu, proto v hierarchii zauímají jednu z nejvyšších pozic (Obr. 4).



Obr. 4: Znalostní hierarchie

Podle Obr. 4 je patrné, že šum zaujímá nejnižší pozici [PK99]. Obsahuje znalostní položky bez významu pro uživatele či systém. Na vyšší úrovni jsou data. Mají jen potenciální význam, protože data jsou znehodnocena nedůležitým šumem. Po zpracování dat vznikají informace. Informace obsahují oddělené významné položky. Znalosti jsou specializované informace (jsou zaměřené na určitou oblast). Metaznalosti jsou znalosti o znalostech. V multiznalostních expertních systémech (ES určený pro více specifických oblastí) se metaznalosti používají k rozlišení znalostí, které budou použity k řešení daného problému.

Funkci znalostí dat v procesu rozhodování znázorňuje Obr. 5



Obr. 5: Proces získávání dat a znalostí

Většinou je zdroj znalostí (expert) odlišný od zdroje informací (shromážděná data). Rozlišení dat a znalostí je následující [Bkol98]:

- Když se můžeme spolehnout při sběru dat na automatický proces, mluvíme o datech.
- Když hledáme experta, který poskytne informace, mluvíme o znalostech.

3.1.1 Logika

Znalosti lze reprezentovat symboly logiky [PK99]. Nejstarší logika pochází od řeckého filozofa Aristotela (4. století př.n.l.). Aristotelova logika vycházela ze sylogismu. Sylogismus má dva předpoklady a jeden výrok. Ten je vyvozen z předpokladů. Klasický příklad sylogismu je:

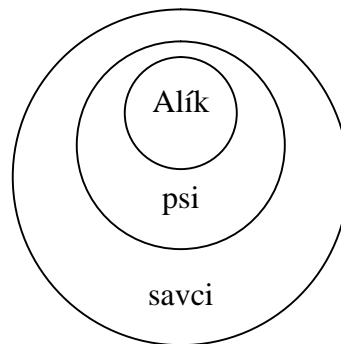
předpoklad: Všichni psi jsou savci.

předpoklad: Alík je pes.

výrok: Alík je savec.

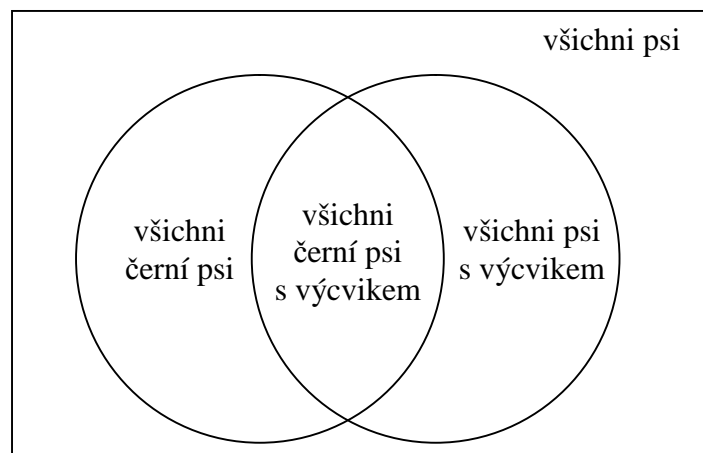
Další formou vyjádření znalostí jsou Vennovy diagramy. Na Obr. 6 je zobrazen stejný problém, kterým se zabýval minulý sylogismus. Vnější kruh vyjadřuje všechny savce, prostřední všechny psi a vnitřní našeho Alíka. Alík je obsažen

v prostředním kruhu a tím pádem je pes. Psi jsou obsaženi v kruhu savci. Z toho plyne, že psi jsou savci. Jelikož je Alík pes a psi jsou savci, tak Alík je savec.



Obr. 6: Vennův diagram

Vennovy diagramy znázorňují množiny objektů. Je tedy možné na ně použít operace z teorie množin (Obr. 7).



Obr. 7: Vennův diagram s konjunkcí

3.1.2 Deduktivní logika

Deduktivní logika je typická metoda k dokazování platnosti tvrzení [PK99]. Tvrzení je souhrn údajů. Poslední údaj má být dokázán na základě předchozích údajů v inferenci.

Zkrácený zápis sylogismu:

předpoklad: Všichni psi jsou savci.

předpoklad: Alík je pes.

výrok: Alík je savec.

je:

Všichni psi jsou savci.

Alík je pes.

∴ Alík je savec

Znak ∴ znamená „proto“.

Tvrzení lze zapsat pomocí pravidel KDYŽ...POTOM

KDYŽ: Všichni psi jsou savci.

Alík je pes.

POTOM: Alík je savec.

3.1.3 Výroková logika

Použití sylogismů a deduktivní logiky nemusí při inferenci stačit, protože pokrývají relativně malé části všech možných logických tvrzení. Možným řešením je použití výrokové logiky a logických proměnných [PK99].

Tvrzení může být ve výrokové logice vyjádřeno:

Jestliže je baterie nabitá, pak svítilna bude svítit.

Baterie je nabitá.

∴ Svítilna bude svítit.

Pomocí logických proměnných může být přepsáno

p = Baterie je nabitá.

q = Svítilna bude svítit.

do výrokové formy:

$p \rightarrow q$

p

∴ q

Pro inferenční mechanismy expertních systémů je důležitá výroková logika. Základními elementy výrokové logiky jsou atomy. Z nich se tvoří výrokové formule pomocí negace, disjunkce, implikace či ekvivalence (Tab. 2).

logická spojka	význam
\sim	negace
\wedge	konjunkce (a)
\vee	disjunkce (nebo)
\rightarrow	implikace (když ... potom)
\leftrightarrow	ekvivalence (tehdy a jen tehdy)

Tab. 2: Logické spojení ve výrokové logice

3.2 PREDIKÁTOVÁ LOGIKA

Predikátová logika je velmi dobře prozkoumaný systém pro reprezentaci a zpracování znalostí [Bkol98]. Původně byla zkoumána pro potřeby matematiky. Ale pro její univerzálnost umožňující analyzovat jazykové výrazy se uplatnila v reprezentaci znalostí. Je vhodná pro studium odvoditelnosti (inference) mezi tvrzeními. Predikátová logika prvního řádu je základem symbolických programovacích jazyků (Prolog).

Jazyk predikátové logiky obsahuje:

- individuové proměnné,
- predikátové symboly,
- kvantifikátory,
- logické spojky,
- funkční symboly a konstanty.

Konkrétní jazyk vždy obsahuje alespoň jeden predikátový symbol, konstanty a funkční symboly nejsou povinné [PK99]. Jsou však dobrou pomůckou ke zlepšení srozumitelnosti a efektivní odvozování. Funkční symboly slouží k vytváření formulí

(jazykové výrazy, které reprezentují znalosti). Důležité jsou také sentace. Sentace jsou formule, ke kterým lze přiřadit pravdivostní hodnotu (chovají se jako výroky).

forma	schéma	predikátová prezentace
A	pro všechna x, všechna S jsou P	$(\forall x) (S(x) \rightarrow P(x))$
E	pro všechna x, žádná S nejsou P	$(\forall x) (S(x) \rightarrow \sim P(x))$
I	pro některá x, x jsou S a P	$(\exists x) (S(x) \wedge P(x))$
O	pro některá x, x jsou S a nejsou P	$(\exists x) (S(x) \wedge \sim P(x))$

Tab. 3: Reprezentace kategorických tvrzení predikátovou logikou

Ze sémantického hlediska mají nejdůležitější váhu pojmy splnitelnost a interpretace. V predikátové logice platí, že co je dokazatelné je i pravdivé, a co je pravdivé je i dokazatelné.

Z výpočtového hlediska je důležité, že je možné logickou dokazatelnost automatizovat.

schéma	predikátová prezentace
pro všechna x, x jsou P	$(\forall x) P(x)$
některá x jsou P	$(\exists x) P(x)$
ne všechna x jsou P nebo některá x jsou P	$\sim(\forall x) P(x)$
žádné x není P	$(\forall x) \sim P(x)$
pro všechna x existuje y rovné P	$(\forall x) (\exists y) p(x,y)$
některá x nejsou P	$(\exists x) \sim P(x)$

Tab. 4: Význam dalších elementárních tvrzení

Predikátová logika pracuje s kvantifikátory \forall a \exists [PK99]. Umožňují potvrzovat nebo vyvracet platnost tvrzení ve tvaru:

$$\frac{(\forall x) \phi(x)}{\therefore \phi(e)}$$

kde ϕ je jakýkoliv výrok či výroková funkce a e je instance (případ), která představuje konkrétní fakt. x je proměnná představující jakýkoliv fakt. Konkrétní ukázka faktu, že Alík je savec je znázorněna:

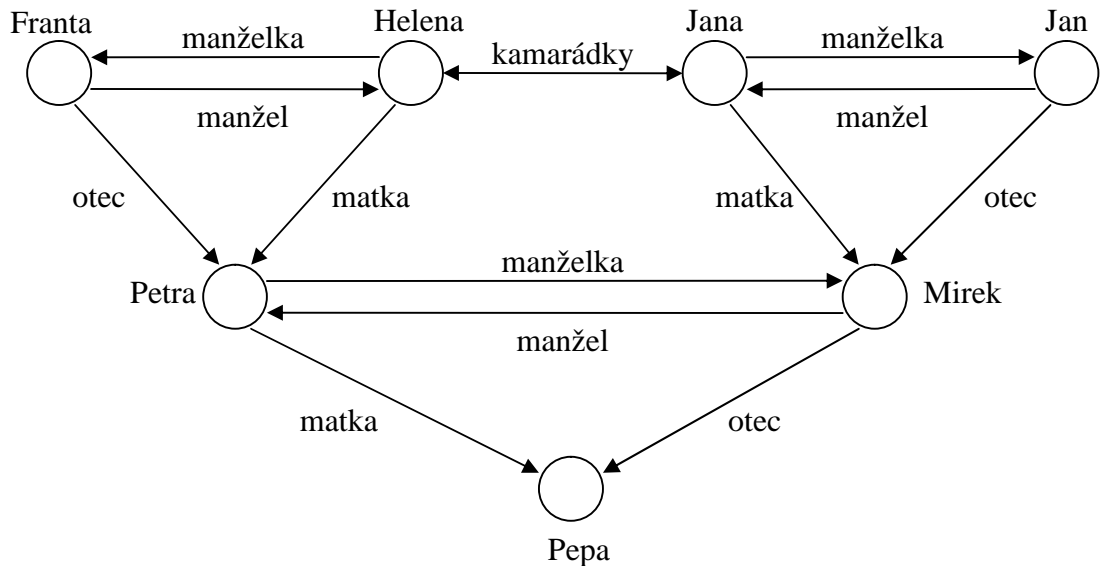
$$\frac{(\forall x) H(x)}{\therefore H(\text{Alík})}$$

kde $H(x)$ je výroková funkce vyjadřující fakt, že x je savec.

3.3 SÉMANTICKÉ SÍTĚ

Sémantické sítě byly vyvinuty v první polovině 60. let minulého století [Bko198]. Měly sloužit k porozumění přirozeného jazyka jako model asociativní paměti člověka. Sémantická síť umožňuje popisovat realitu jako objekty, které jsou navzájem v nějakých vztazích. Sítě jsou tvořeny grafy s uzly představující objekty. Uzly mohou být propojeny hranami představující relace. Relace představují základní prostředek pro vyjádření znalostí. Osvědčily se typické relace jako AS-A (ISA), A-KIND-OF (AKO) a PART-OF. ISA umožňuje, že určitý objekt patří do určité třídy objektů. AKO vyjadřuje hierarchii tříd. PART-OF umožňuje vyjádřit objekt tak, že určité části, ze kterých je tvořen (např. procesor a paměť jsou částí počítače).

Sémantická síť může popisovat závislosti. Lze to demonstrovat třeba na rodinných svazcích (Obr. 8).



Obr. 8: Sémantická síť popisující rodinné vztahy

Výhody sémantických sítí [Dvo04]:

- explicitní a jasné vyjádření,
- snížená doba hledání (pro dotazy typu dědičnosti či rozpoznávání).

Nevýhody:

- neexistence interpretačních standardů,
- neexistence standardních definic jmen vazeb,
- nebezpečí chybné inference,
- nebezpečí kombinatorské exploze.

3.4 RÁMCE

Rámce byly navrženy ve druhé polovině 60. let minulého století [Bko198]. Rámce měly původně umožňovat reprezentaci stereotypních informací. Práce s rámci měla být založena na postupném vyplňování stránek. Do nich se zapisují vlastnosti položek. Často se používá předdefinovaných hodnot.

Výhodou rámce je popis složitějších objektů s mnoha relacemi. Jazyky založené na rámci jsou například KEE a LISP s rámcovým rozšířením (LOOPS). Rámce jsou obdobou ke strukturám či objektům u vyšších programovacích jazyků

(např. Pascal, C). Pole a hodnoty se u rámců nazývají sloty a náplně. Tab. 5 zobrazuje příklad rámce popisující auto.

slot	náplň
výrobce	Ford
model	Escort
rok výroby	1988
objem motoru	1,6
motor	benzinový
počet dveří	3
barva	šedá

Tab. 5: Rámec popisující auto

Rámec zobrazený v Tab. 5 je velmi jednoduchý. Ve složitějších aplikacích mohou sloty obsahovat poznámky, grafické objekty, pravidla, otázky, poznámky, hypotézy, procedury a vnořené rámce. Obecně se v rámcích vyskytují tři typy procedur. Jsou to:

- if-needed – vykoná se, když není k dispozici inicializační hodnota a je požadována náplň slotu,
- if-added – vykoná se, když uživatel vkládá náplň slotu,
- if-removal – procedura se spustí, když uživatel odstraňuje náplň slotu.

Rámce bývají obvykle doplněny o pravidla, která dovolují odvozovat v bázi znalostí.

Styl programování využívající objekty se jmenuje objektově orientované programování (OOP). Základní vlastnosti OOP jsou [Bko198]:

- zapouzdření – spojení definice datových struktur s procedurami do jednoho datového typu (objektu),

- dědičnost – objekty na nižší úrovni dědí definice datových typů a metod od objektů na vyšší úrovni,
- polymorfismus – metoda, která se dědí se stejným jménem, může mít na každé úrovni jinou podobu.

Systémy využívající rámce patří například Nexpert Object, PC-Kappa nebo CLOS.

Výhody rámců [Dvo04]:

- snazší usuzování řízené očekáváním,
- organizace znalostí,
- samořízení,
- uchování dynamických hodnot (ve slotech).

Možné nevýhody:

- potíže s odlišností objektů od prototypu,
- obtížné přizpůsobení novým situacím,
- obtížný popis detailních heuristických znalostí.

3.5 PRAVIDLA

Nejčastěji používané expertní systémy jsou založené na pravidlech [Bko198]. Pravidlovým systémům se také říká produkční systémy. Využití pravidlových expertních systémů je výhodné zejména pro:

- modularitu – jednoduché a jednoznačné vyjádření určité znalosti s možností rozšíření,
- vysvětlování – jednoduché vytvoření vysvětlovacího modulu,
- podobnost s usuzováním člověka – pravidla typu „když...potom“ přibližně odpovídají usuzování člověka a je snadné vytvořit pravidla na základě expertovy zkušenosti.

Znalosti v pravidlových systémech jsou reprezentovány pomocí pravidel, která mohou mít například takovéto tvary:

IF předpoklad THEN závěr

IF situace THEN akce

IF podmínka THEN závěr AND akce

IF podmínka THEN důsledek1 ELSE důsledek2

Část pravidla za IF (levá strana pravidla) se nazývá antecedent (podmínková část) nebo také část vzorů [Dvo04]. Tato část může být složena z individuálních podmínek nebo vzorů. Část pravidla za THEN (pravá strana pravidla) se nazývá konsekvent a může také obsahovat několik akcí nebo závěrů. V předpokladové části se mohou vyskytnout spojky AND a OR, v důsledkové části se může vyskytnout spojka AND. Součástí pravidla může být také tzv. kontext, ve kterém má být pravidlo uvažováno.

Příklady pravidel:

IF auto_startuje = ne AND světla_svíí = ne THEN diagnóza = vybitá_baterie

IF chování_dítěte = dobré THEN možnost_uhlí = nízká AND
možnost_sladkostí = vysoká

Pravidlo může také obsahovat neurčitosti.

IF výška_stromu(X) = malá THEN šířka_stromu(X) = malá

Výhody pravidlových systémů:

- modularita,
- uniformita,
- přirozenost.

Možné nevýhody a problémy pravidlových systémů:

- nebezpečí nekonečného řetězení,
- neefektivnost,
- neprůhlednost,
- přidání nové znalosti,
- modifikace pravidel,
- pokrytí domény (domény mohou vyžadovat velké množství pravidel).

4. EXPERTNÍ SYSTÉM NPS32

Expertní systém NPS vznikl v rámci diplomové práce Pavla Vojty ve školním roce 2000/2001. Rok nato přidal Petr Skorkovský multimediální podporu a propojení s externími programy.

4.1 MATEMATICKÝ APARÁT

NPS32 je založen na práci s pravděpodobností [Jir08]. Pravděpodobnosti jsou vyjádřené pomocí dvou hodnot (T, F). Zavedením těchto dvou hodnot zjednodušilo výpočet vztahů a zároveň umožňuje zavedení informace o spornosti některého tvrzení.

Výraz (1-T) vyjadřuje míru důvěry v pravdivosti tvrzení a naopak výraz (1-F) míru důvěry v nepravdivost tvrzení. Z toho vyplývá:

(0, 1) odpovídá pravdivému tvrzení,

(1, 0) odpovídá nepravdivému tvrzení,

(1, 1) znamená nepřítomnost jakékoli informace,

(0, 0) znamená spor v bázi znalostí nebo v odpovědích uživatele.

Pro výpočty s hodnotami T a F platí vztahy:

$$1. \quad T = \frac{F(1-p)}{p} \quad (4.1)$$

$$2. \quad F = \frac{pT}{1-p} \quad (4.2)$$

kde p je pravděpodobnost. (Pro T, F \in <0, 1> je p \in <0, 1>)

Jestliže je p<0,5 (50%) volíme maximální T (T=1) a F dopočteme pomocí vztahu (4.2). Pokud je p>0,5 volíme maximální F (F=1) a T dopočteme pomocí vztahu (4.1).

Pravděpodobnost uzlu je dána:

$$P = p \cdot 100 = \frac{F}{T+F} \cdot 100 \text{ [%]} \quad (4.3)$$

4.2 BÁZE ZNALOSTÍ

Báze znalostí nabývá dvou podob [Jir08]. První je soubor uložený v textovém souboru. Formát textové podoby je založen na expertním systému DrExpert. Textový zápis poskytuje úplnou informaci o bázi znalostí, jejích prvcích a vzájemných vztazích mezi nimi.

Druhá podoba nabývá báze, když je nahrána do NPS32. Hlavní prvky báze (uzly) jsou reprezentovány samostatnými paměťovými strukturami. Ty jsou v programu identifikovány svou adresou, ne jménem. Podobně jsou reprezentovány i vazby. Díky tomu je například přejmenování uzlu či přidání vazby nenáročné.

Bázi lze rozdělit na dvě části. První část obsahuje informace o celé bázi. Druhá část definuje uzly. První část může obsahovat:

- Název báze – libovolný název, může se lišit od názvu souboru. V textové podobě báze je uvozen znaky „;“.
- Informační text k bázi – slouží k popisu báze. Popis je uvozen znaky „;“.
- Počet implicitně nabízených odpovědí – definuje počet možných odpovědí, které systém nabídne, pokud není jinak definováno v uzlu. Možné hodnoty jsou 3, 5 a 7. Hodnota je v bázi uvozena znaky „;?“.

Za úvodní částí následují definice uzlů v hlavní části báze znalostí. Definice uzlů jsou uspořádány v logickém sledu. Uzel musí být tedy první definován a až pak se na něj může odkazovat.

V bázi jsou tři druhy uzlů – dotazovatelné, pomocné a cílové. Dotazovatelné uzly se používají pro reprezentaci otázky v bázi znalostí. Dělí se na dvě skupiny:

- Přímý dotazovatelný uzel – dotazy se pokládají v pořadí v jakém jsou definovány v bázi znalostí. Tento uzel je označen znakem „D“.
- Běžně dotazovatelný uzel – inferenční mechanismus pokládá otázky definované těmito uzly podle svého uvážení. Uzel je označen znakem „A“.

Přímý dotazovatelný uzel může být kvantitativního typu. Je označen přidáním znaku „K“, celý uzel je tedy značen znaky „DK“. Tento uzel musí obsahovat definici uživatelských odpovědí.

Pomocné uzly jsou nejjednoduššími typy uzlů. Jeho definice je dána jeho identifikátorem a počáteční pravděpodobností. Neobsahuje žádný parametr. Používá se jen ve dvou případech. V prvním se používá jako pomocný neviditelný uzel. Je potvrzen či zavržen hypotézami zastoupenými dotazovatelnými a pomocnými uzly.

Ve druhém případě se používá ve spojení s kvantitativním uzlem. Pomocný uzel zastupuje hypotézu reprezentovanou odpovědí kvantitativního uzlu.

Cílové uzly zastupují hypotézy, které systém vybere jako nejpravděpodobnější. Textový zápis definice uzlu se skládá z identifikátoru, počáteční pravděpodobnosti, parametru identifikující cílový uzel „G“ a řádku s názvem hypotézy.

Výsledkem expertízy je seznam cílových uzlů seřazených sestupně, od nejpravděpodobnějšího po nejnepravděpodobnější. Báze znalostí by měla být navržena tak, že by se nejpravděpodobnější cílová hypotéza blížila 100 %. Nevyhovující hypotéza by se měla blížit k nule. Hodnota 50 % znamená nedostatek informací nebo spor.

5. SROVNÁNÍ RZ

Samotné srovnání metod reprezentace znalostí předvedu na expertíze výběru vhodné grafické karty (GK) do nového PC. Grafické karty mají mnoho podstatných i nepodstatných parametrů, které jsou přímo ideální pro expertízu.

Zaměřím se hlavně na efektivitu zápisu a její modifikovatelnost.

5.1 RÁMCE

Pro začátek si ukážeme jednoduchý zápis pro 4 grafické karty s malým počtem parametrů (Tab. 6).

parametr/typ	HD 3870 X2	HD 3650	9800GX2	8600GT
značka	ATI	ATI	NVIDIA	NVIDIA
účel	hry	multimédia	hry	multimédia

Tab. 6: Rámce čtyř grafických karet – jednoduchý zápis

Rámec zobrazuje jen značku grafické karty a její účel. Zápis báze znalostí v textové podobě (ramce1.n32) vypadá takto:

- (1) ;.Výběr vhodné grafické karty do PCIE. Rámce1.n32 – jednoduchý zápis.
- (2) ;?5
- (3) .HRY $p=50\%$
- (4) .MEDIA $p=50\%$
- (5) .UCEL $p=50\%$ DK
- (6) ;K čemu bude grafická karta primárně sloužit?
- (7) ;*Hraní PC her [1]*
- (8) ;*Multimédia (PC určené k přehrávání audia a videa)[2]*
- (9) HRY
- (10) MEDIA
- (11) .ATI $p=50\%$ D
- (12) ;Preferujete značku ATI?
- (13) ;(Jestli preferujete Nvidii, tak zvolte odpověď - ne.)
- (14) .A3650S1 $p=50\%$ G
- (15) ;ATI HD 3650 Sapphire
- (16) & 90% +ATI
- (17) & 50% -HRY
- (18) & 85% +MEDIA
- (19) .A3870X2P $p=50\%$ G
- (20) ;ATI HD 3870 X2 Powercolor
- (21) & 90% +ATI

- (22) & 75% +HRY
- (23) & 75% -MEDIA
- (24) .N86GTII p=50% G
- (25) ;NVIDIA 8600GT Inno3D
- (26) & 90% -ATI
- (27) & 60% -HRY
- (28) & 85% +MEDIA
- (29) .N98X2A2 p=50% G
- (30) ;NVIDIA 9800X2 Asus
- (31) & 90% -ATI
- (32) & 80% +HRY
- (33) & 75% -MEDIA
- (34) #

Zápis báze znalostí je velmi krátký a snadný, protože má jen čtyři cílové uzly a dva dotazovatelné. Systém se pouze ptá uživatele k jakému účelu bude grafická karta sloužit (6. řádek). Buď bude sloužit na hraní her nebo bude umístěná v PC učeném pro přehrávání filmů a hudby. Další otázka (12. řádek) se ptá uživatele jestli preferuje určitou značku. Cílové uzly s údaji pro přepočet báze znalostí začínají od 14. řádku.

parametr/typ	HD 3870 X2	HD 3870	HD 3650	HD 3450
značka	ATI	ATI	ATI	ATI
účel	hry	hry	multimédia	multimédia
výkon	velmi vysoký	vysoký	střední	nízký
cena [Kč]	8000,-	3300,-	1300,-	730,-
parametr/typ	9800GX2	9600GT	8600GT	8400GS
značka	NVIDIA	NVIDIA	NVIDIA	NVIDIA
účel	hry	hry	multimédia	multimédia
výkon	velmi vysoký	vysoký	střední	nízký
cena [Kč]	11500,-	3300,-	1500,-	670,-

Tab. 7: Rámce osmi grafických karet – rozšířený zápis

Tahle expertíza je však nedostatečná. Čtyři cílové uzly je velmi málo, když je jich na trhu několik stovek. Taky chybí podstatná otázka, která zajímá většinu zákazníků – cena. Proto si zápis ve formě rámců trochu rozšíříme (Tab. 7).

Tento zápis rozšiřuje předchozí zápis o čtyři cílové uzly a zdvojnásobil počet parametrů jednotlivých grafických karet. Textový zápis báze znalostí se z původních 34 řádků zvětšil na 121 řádků. Celý zápis báze znalostí (ramce2.n32) není nutné vypisovat. Uvedu pouze nejpodstatnější změny:

- (1) ;.Výběr vhodné grafické karty do PCIE. Rámce2.n32 – rozšířený zápis.
- (2) ;?5
- (3) .HRY $p=50\%$
- (4) .MEDIA $p=50\%$
- (5) .UCEL $p=50\%$ DK
- (6) ;K čemu bude grafická karta primárně sloužit?
- (7) ;*Hraní PC her [1]*
- (8) ;*Multimédia (PC určené k přehrávání audia a videa)[2]*
- (9) HRY
- (10) MEDIA
- (11) .VYKON $p=50\%$ D
- (12) ;Záleží Vám na výkonu?
- (13) .DO800 $p=50\%$
- (14) .DO1500 $p=50\%$
- (15) .DO3300 $p=50\%$
- (16) .DO8000 $p=50\%$
- (17) .NAD8000 $p=50\%$
- (18) .CENA $p=50\%$ DK
- (19) ;V jaké cenové relaci (s DPH) se má karta pohybovat?
- (20) ;*Do 800,- [1]*
- (21) ;*Do 1 500,- [2]*
- (22) ;*Do 3 300,- [3]*
- (23) ;*Do 8 000,- [4]*
- (24) ;*Nad 8 000,- [5]*
- (25) DO800
- (26) DO1500
- (27) DO3300
- (28) DO8000
- (29) NAD8000
- (30) .ATI $p=50\%$ D
- (31) ;Preferujete značku ATI?
- (32) ;(Jestli preferujete Nvidii, tak zvolte odpověď - ne.)

Jak je patrné, přibyly dva dotazovatelné uzly na řádcích 11 až 29. Ostatní dotazovatelné uzly jsou identické jako v předchozím příkladu. Pro předvedení změn cílových uzlů stačí prezentovat pouze jeden cílový uzel (Tab. 8). Přibyly pouze

parametry k přepočítání báze znalostí na řádcích 49 až 54. Hodnoty zvolených parametrů není potřeba komentovat, protože nejsou nijak zvlášť důležité pro zhodnocení samotných rámců.

Cílový uzel s jednoduchého zápisu	Cílový uzel z rozšířeného zápisu
(14) <i>.A3650S1 p=50% G</i>	(44) <i>.A3650S1 p=50% G</i>
(15) <i>;ATI HD 3650 Sapphire</i>	(45) <i>;ATI HD 3650 Sapphire</i>
(16) <i>& 90% +ATI</i>	(46) <i>& 90% +ATI</i>
(17) <i>& 50% -HRY</i>	(47) <i>& 50% -HRY</i>
(18) <i>& 85% +MEDIA</i>	(48) <i>& 85% +MEDIA</i>
	(49) <i>/ 60% -VYKON</i>
	(50) <i>& 90% -DO800</i>
	(51) <i>& 90% +DO1500</i>
	(52) <i>/ 60% -DO3300</i>
	(53) <i>/ 60% -DO8000</i>
	(54) <i>/ 60% -NAD8000</i>

Tab. 8: Ukázka změn cílového uzlu u rámců

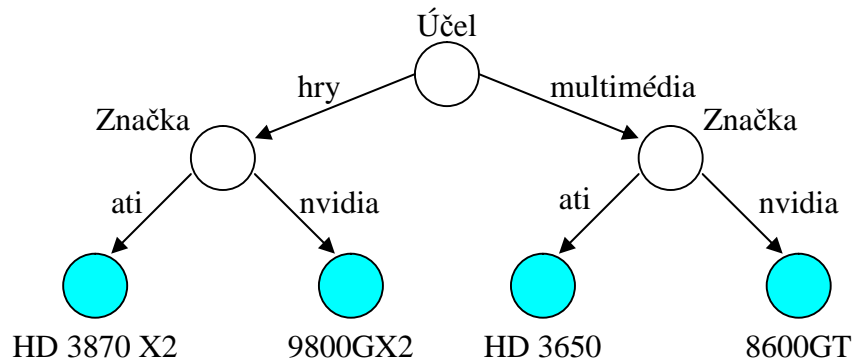
5.1.1 Zhodnocení

Jak je vidět z uvedeného příkladu, rámce jsou ideální metodou reprezentace znalostí v expertních systémech pro úlohy typu výběr vhodného výrobku, služby či nějaké jiné věci (cílového uzlu) o parametrech, které se dají kategorizovat. Takto zobrazené cílové uzly jsou snadno modifikovatelné a přehledné i s přibývajícimi cílovými, dotazovatelnými a pomocnými uzly.

Nevýhodou je zbytečné vypisování parametrů pro přepočet báze znalostí v cílových uzlech. Na první pohled nejde určit, zda jsou tyto parametry podstatné či nikoli.

5.2 SÉMANTICKÉ SÍTĚ

Rámce z Tab. 6 se dají pomocí sémantické sítě zapsat do podoby na Obr. 9. Pro tak malý počet dotazovatelných a cílových uzlů je to dokonce přehlednější než rámce.



Obr. 9: Jednoduchá sémantická síť

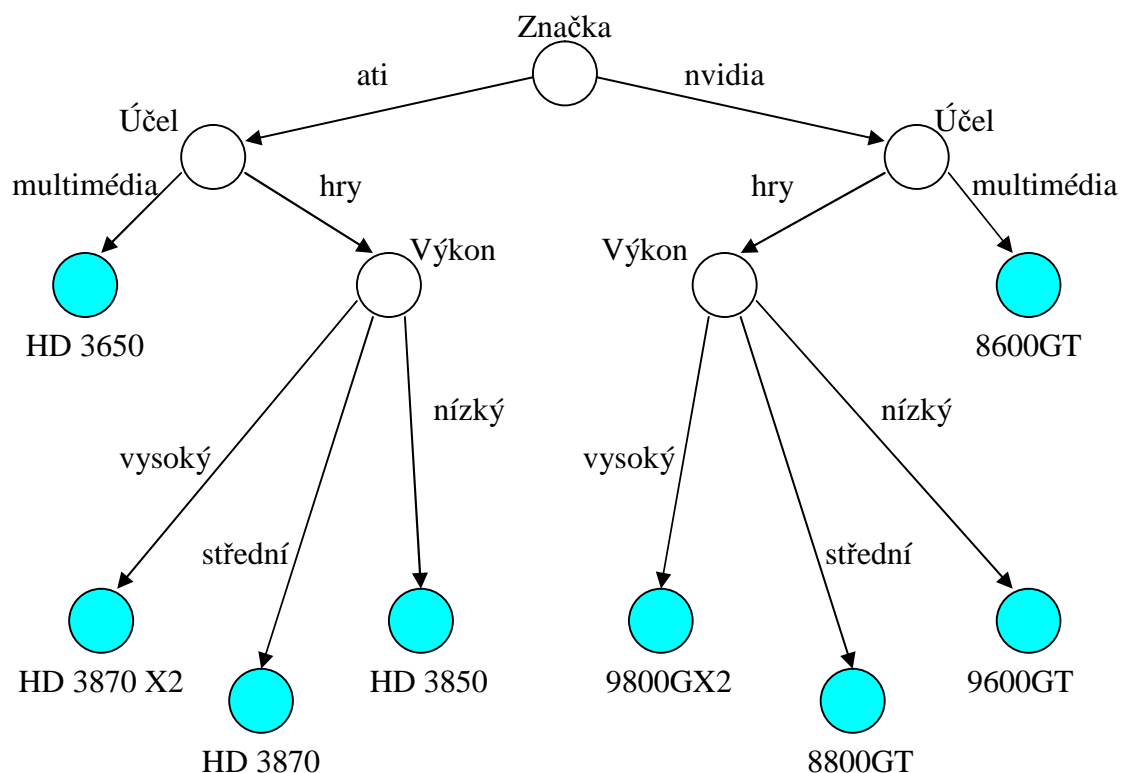
Na začátku je první dotazovatelný uzel, který se ptá na účel grafické karty. Z tohoto dotazovatelného uzlu vedou dvě vazby (hry, multimédia) na další dotazovatelné uzly Značka. Z těchto uzlů vedou vazby na cílové uzly.

Zápis textové podoby báze znalostí by mohla být identická jako v rámci – jednoduchý zápis. Ale aby zápis odpovídal síti je potřeba udělat pár změn (sit1.n32):

- (1) ;.Výběr vhodné grafické karty do PCIE. Sit1.n32 – jednoduchý zápis.
- (2) .HRY $p=50\%$
- (3) .MEDIA $p=50\%$
- (4) .UCEL $p=50\%$ DK
- (5) ;K čemu bude grafická karta primárně sloužit?
- (6) ;*Hraní PC her [1]*
- (7) ;*Multimédia (PC určené k přehrávání audia a videa)[2]*
- (8) HRY
- (9) MEDIA
- (10) .ATI $p=50\%$
- (11) .NVIDIA $p=50\%$
- (12) .ZNACKA $p=50\%$ DK
- (13) ;Jakou značku preferujete?
- (14) ;*ATI [1]*
- (15) ;*NVIDIA [2]*
- (16) ATI
- (17) NVIDIA
- (18) .A3650S1 $p=50\%$ G
- (19) ;ATI HD 3650 Sapphire
- (20) & 90% +ATI
- (21) & 90% -NVIDIA
- (22) & 90% -HRY
- (23) & 90% +MEDIA
- (24) .A3870X2P $p=50\%$ G
- (25) ;ATI HD 3870 X2 Powercolor
- (26) & 90% +ATI

- (27) & 90% -NVIDIA
- (28) & 90% +HRY
- (29) & 90% -MEDIA
- (30) .N86GT11 p=50% G
- (31) ;NVIDIA 8600GT Inno3D
- (32) & 90% -ATI
- (33) & 90% +NVIDIA
- (34) & 90% -HRY
- (35) & 90% +MEDIA
- (36) .N98X2A2 p=50% G
- (37) ;NVIDIA 9800X2 Asus
- (38) & 90% -ATI
- (39) & 90% +NVIDIA
- (40) & 90% +HRY
- (41) & 90% -MEDIA
- (42) #

Z textového zápisu je patrné, že jsou teď oba dotazovatelé uzly kvantitativního typu. Na konci expertízy budou mít cílové uzly vždy 99, 50, 50 a 1%, aby výsledek expertízy odpovídal sémantické síti (Obr. 9). Dalo by de to udělat i tak, aby cílové uzly vyšly 99, 1, 1 a 1%. Ale zápis by byl o něco složitější. Tato síť je však velmi jednoduchá, proto ji rozšíříme (Obr. 10).



Obr. 10: Rozšířená sémantická síť

Do nové sémantické sítě jsem přidal jen čtyři cílové uzly a jeden dotazovatelný uzal. I po takto malé změně se sémantická síť značně rozrostla. Zatím je ještě docela přehledná, ale kdyby jsem přidal více uzlů, vznikl by z toho extrémní moloch.

Otázky v textové bázi znalostí (sit2.n32) vypadají takhle:

- (1) ;.Výběr vhodné grafické karty do PCIE. Sit2.n32 – rozšířený zápis.
- (2) .ATI p=50%
- (3) .NVIDIA p=50%
- (4) .ZNACKA p=50% DK
- (5) ;Jakou znacku preferujete?
- (6) ;*ATI [1]*
- (7) ;*NVIDIA [2]*
- (8) ATI
- (9) NVIDIA
- (10) .HRY p=50%
- (11) .MEDIA p=50%
- (12) .UCEL p=50% DK
- (13) ;K čemu bude grafická karta primárně sloužit?
- (14) ;*Hraní PC her [1]*
- (15) ;*Multimédia (PC určené k přehrávání audia a videa)[2]*
- (16) HRY
- (17) MEDIA
- (18) .NIZKY p=50%
- (19) .STREDNI p=50%
- (20) .VYSOKY p=50%
- (21) .VYKON p=50% DK
- (22) ;Jaký má mít GK výkon?
- (23) ;*Nízký [1]*
- (24) ;*Střední [2]*
- (25) ;*Vysoký [3]*
- (26) NIZKY
- (27) STREDNI
- (28) VYSOKY
- (29) < 51% MEDIA

Na řádcích 2 až 17 jsou dva dotazovatelné uzly kvantitativního typu. Jestli se zvolí v druhé otázce (13. řádek) odpověď – hraní her, tak se položí třetí otázka (22. řádek).

Z ukázky (Tab. 9) je patrné, že první uzal zůstal beze změn. U druhého uzlu přibýly parametry týkající se výkonu, které přibýly v sémantické síti (Obr. 10). Za to zmizel parametr „& 90% +HRY“, které nahrazují právě výkonnostní parametry.

Cílové uzly s jednoduchého zápisu	Cílové uzly z rozšířeného zápisu
(18) <i>.A3650S1 p=50% G</i> (19) <i>;ATI HD 3650 Sapphire</i> (20) <i>& 90% +ATI</i> (21) <i>& 90% -NVIDIA</i> (22) <i>& 90% -HRY</i> (23) <i>& 90% +MEDIA</i>	(30) <i>.A3650S1 p=50% G</i> (31) <i>;ATI HD 3650 Sapphire</i> (32) <i>& 90% +ATI</i> (33) <i>& 90% -NVIDIA</i> (34) <i>& 90% -HRY</i> (35) <i>& 90% +MEDIA</i>
(24) <i>.A3870X2P p=50% G</i> (25) <i>;ATI HD 3870 X2 Powercolor</i> (26) <i>& 90% +ATI</i> (27) <i>& 90% -NVIDIA</i> (28) <i>& 90% +HRY</i> (29) <i>& 90% -MEDIA</i>	(52) <i>.A3870X2P p=50% G</i> (53) <i>;ATI HD 3870 X2 Powercolor</i> (54) <i>& 90% +ATI</i> (55) <i>& 90% -NVIDIA</i> (56) <i>& 90% -MEDIA</i> (57) <i>& 90% -NIZKY</i> (58) <i>& 90% -STREDNI</i> (59) <i>& 90% +VYSOKY</i>

Tab. 9: Ukázka změn cílových uzlů u sémantických sítí

5.2.1 Zhodnocení

Sémantická síť je efektivní ve vyhledávání cílového uzlu. Pokud máte k dispozici velké plátno, tak je i snadno rozšiřitelná. Při psaní textové verze báze znalostí se podle ní snadno orientuje a nevznikají irelevantní parametry.

Nevýhodou může být nepřehlednost při velkém množství uzlů a vazeb mezi nimi. Taky není vhodné používat odpovědi typu „nevím“. Tyto situace zavedené do sémantické sítě ji velmi komplikují.

5.3 PRAVIDLA

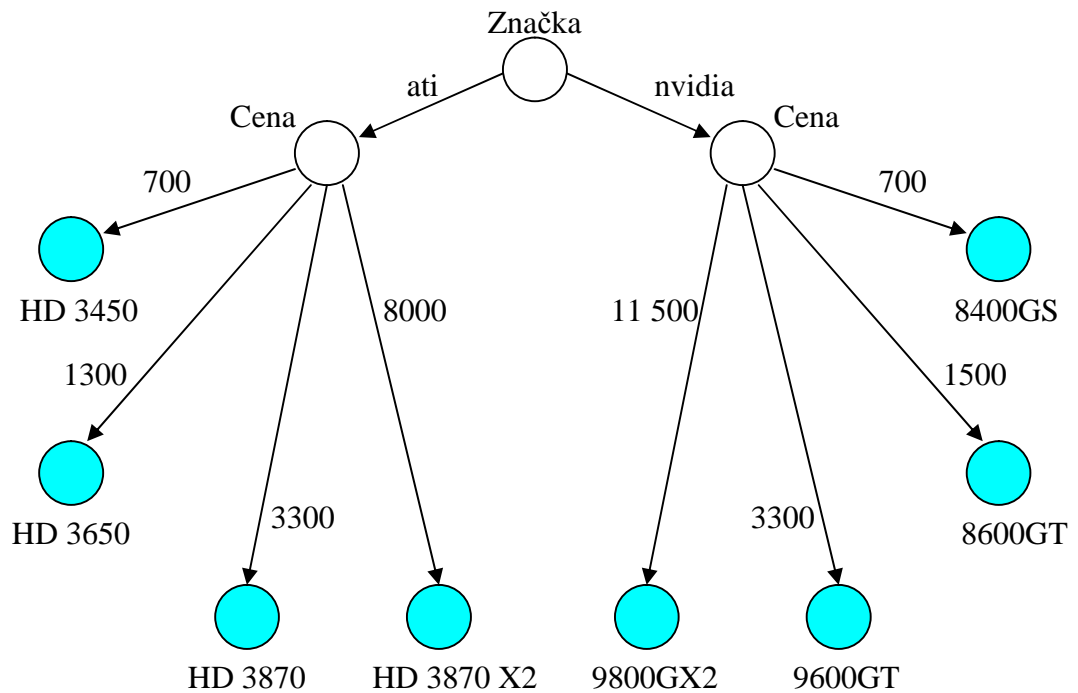
Zápis sémantické sítě (Obr. 9) se dá pomocí pravidel zapsat:

```

IF značka=ati      AND účel=multimédia      THEN gk=hd3650
IF značka=ati      AND účel=hry              THEN gk=hd3870x2
IF značka=nvidia   AND účel=multimédia      THEN gk=8600GT
IF značka=nvidia   AND účel=hry              THEN gk=9800gx2
  
```

Zápis textové formy báze znalostí vypadá identicky jako u sémantické sítě (sit1.n32). Zápis je možné trochu pozměnit, aby víc odpovídal zápisu pomocí pravidel (pravidl1.n32 – viz. příloha). Tento zápis odpovídá i sémantické

síti (Obr. 9). Rozšiřitelnost této metody je vhodná při malém počtu atributů a mnoha parametrech. Nejsnadněji to vyjádří sémantická síť (Obr. 11).



Obr. 11: Vyjádření rozšíření pravidel pomocí sémantické sítě

Sémantická síť (Obr. 11) je pomocí pravidel zapsána takto:

IF značka=ati	AND cena=700	THEN gk=hd3450
IF značka=ati	AND cena=1300	THEN gk=hd3650
IF značka=ati	AND cena=3300	THEN gk=hd3870
IF značka=ati	AND cena=8000	THEN gk=hd3870x2
IF značka=nvidia	AND cena =700	THEN gk=8400GS
IF značka=nvidia	AND cena =1500	THEN gk=8600gt
IF značka=nvidia	AND cena =3300	THEN gk=9600gt
IF značka=nvidia	AND cena =11500	THEN gk=9800gx2

Dotazovatelný uzal – Účel – jsem vynechal úmyslně, aby byl zápis přehlednější. Kdybych ho nevynechal, zápis by vypadal takto:

IF značka=ati AND cena=800 AND účel=multimédia THEN gk=hd3450

Ukázka zápisu textové podoby báze znalostí (pravidl2.N32) je v příloze.

5.3.1 Zhodnocení

Pravidla jsou vhodná pro velké množství cílových uzlů o malém počtu atributů, které však mohou nabývat mnoha různých parametrů. Se zvyšujícím se počtem atributů se ztrácí na přehlednosti. Do báze lze snadno připisovat nová pravidla, které se mohou psát na přeskáčku. Pro přehlednost je však při psaní pravidel vhodné dodržovat určitý formát zápisu.

Cílové uzly o velkém počtu atributů jsou složité pro zápis, zvyšuje se totiž počet pravidel. Se zvyšujícím se počtem pravidel se zvyšuje nepřehlednost a neefektivita zápisu.

5.4 PREDIKÁTOVÁ LOGIKA

Obdobný zápis rámců (Tab. 6) lze zapsat pomocí predikátové logiky:

$$\forall X(\text{multimediální}(X) \ \& \ \text{ati}(X) \rightarrow \text{hd3650}(X))$$

$$\forall X(\text{herní}(X) \ \& \ \text{ati}(X) \rightarrow \text{hd3870x2}(X))$$

$$\forall X(\text{multimediální}(X) \ \& \ \text{nvidia}(X) \rightarrow \text{8600gt}(X))$$

$$\forall X(\text{herní}(X) \ \& \ \text{nvidia}(X) \rightarrow \text{9800gx2}(X))$$

Výraz $\forall X(\text{multimediální}(X) \ \& \ \text{ati}(X) \rightarrow \text{hd3650}(X))$ znamená: „Všechny grafiky vhodné pro multimédia se značkou ATI jsou HD 3650.“. Textový zápis báze znalostí je identický jako u předcházející metody (pravidl1.n32).

5.4.1 Zhodnocení

Predikátová logika se moc nehodí na expertízu výběru vhodné grafické karty. Spíše se využívá v odvozování platnosti tvrzení jako je například toto: „Pokud hmyz létá nízko nad zemí, bude pršet. Vlaštky létají nízko nad zemí tehdy, když bude pršet a bude pršet, když budou vlaštky létat nízko.“

Definice logických proměnných:

V = vlaštky létají nízko nad zemí,

H = hmyz létá nízko nad zemí,

P = bude pršet.

Problém se zapíše:

$$H \rightarrow P$$

$$P \leftrightarrow V$$

$$\underline{H}$$

$$\therefore V$$

Premisa $P \leftrightarrow V$ dovoluje provést substituci V za P v premise $H \rightarrow P$.

$$H \rightarrow V$$

$$\underline{H}$$

$$\therefore V$$

Z toho vyplývá, že jestliže bude hmyz létat nízko nad zemí, tak vlaštovky budou také létat nízko nad zemí.

Z názorné ukázky plyne, že predikátová logika je vhodná při popisu výroků a jejich platností. Pomocí logických úprav (například substituce) lze zjednodušit zápis a tím zvýšit jeho efektivitu.

6. ZÁVĚR

V této bakalářské práci byly prostudovány metody reprezentace znalostí v expertních systémech. Nejdříve jsou popsány samotné expertní systémy. Po nich byly charakterizovány jednotlivé metody reprezentace znalostí.

K ukázce porovnání metod posloužil expertní systém NPS32. Primární problematika se týká výběru grafických karet do nového PC. U každé metody byla vytvořena jednoduchá báze znalostí. Tato báze slouží k prezentaci efektivity zápisu metody. Dále byla vytvořena druhá rozšířená báze znalostí, která prezentuje modifikovatelnost metody.

Metoda rámců se hodí k zápisu báze znalostí o mnoha cílových uzlů, které mají velké množství parametrů, které se dají kategorizovat. Rámce jsou snadno modifikovatelné a přehledné. Při ladění báze znalostí, se však ve velkém množství parametrů špatně orientuje a existuje nebezpečí zbytečné aplikace parametru.

Sémantická síť je ideální pro úlohy o malém počtu cílových uzlů a vazeb mezi nimi. Vazby by měli být jen určitého charakteru. S rostoucím počtem cílových uzlů ubývá na přehlednosti.

Pravidla jsou vhodné pro úlohy s velkým počtem cílových uzlů. Tyto uzly by však měli mít co nejmenší počet atributů, které mohou nabývat velkým množstvím hodnot. Modifikovatelnost metody je snadná. I při dodržování určitého stylu zápisu o mnoha pravidlech, se stává zápis nepřehledným.

Predikátová logika slouží k zápisu výroků a jejich platností, které se dají zapsat do formy logiky. Tyto logické formule se pak dají zjednodušit pomocí úprav. Metoda jde lehce modifikovat. Při rozsáhlejších zápisech je metoda nepřehledná.

7. SEZNAM OBRÁZKŮ A TABULEK

Obr. 1: Architektura expertního systému	12
Obr. 2: Architektura diagnostických ES	17
Obr. 3: Architektura plánovacích ES	18
Obr. 4: Znalostní hierarchie	23
Obr. 5: Proces získávání dat a znalostí	24
Obr. 6: Vennův diagram.....	25
Obr. 7: Vennův diagram s konjunkcí	25
Obr. 8: Sémantická síť popisující rodinné vztahy.....	30
Obr. 9: Jednoduchá sémantická síť	41
Obr. 10: Rozšířená sémantická síť	42
Obr. 11: Vyjádření rozšíření pravidel pomocí sémantické sítě.....	45
Tab. 1: Přehled expertních systémů dle oblastí využití	20
Tab. 2: Logické spojení ve výrokové logice	27
Tab. 3: Reprezentace kategorických tvrzení predikátovou logikou.....	28
Tab. 4: Význam dalších elementárních tvrzení.....	28
Tab. 5: Rámec popisující auto.....	31
Tab. 6: Rámce čtyř grafických karet – jednoduchý zápis	37
Tab. 7: Rámce osmi grafických karet – rozšířený zápis	38
Tab. 8: Ukázka změn cílového uzlu u rámců.....	40
Tab. 9: Ukázka změn cílových uzlů u sémantických sítí.....	44

8. POUŽITÁ LITERATURA

- [Bko198] Berka P. a kol.: *Expertní systémy*. Skripta, VŠE Praha, 1998.
- [Cel99] Celbová I.: *Úvod do problematiky expertních systémů*. Ikaros [online]. 1999, roč. 3, č. 8 [cit. 2007-12-20]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/393>>. URN-NBN:cz-ik393. ISSN 1212-5075.
- [Dvo04] Dvořák J.: *Expertní systémy*. Vysoké učení technické, Brno 2004
- [Hab04] Habiballa H.: *Umělá inteligence*, Ostravská Univerzita 2004
- [HR84] Hajičová E., Wichs T.: *Reprezentace znalostí v systémech umělé inteligence*. Skripta, Univerzita Karlova v Praze, 1984
- [Jir08] Jirsík V.: *Přednášky z předmětu "Moderní prostředky v automatizaci"*. ÚAMT FEKT VUT, Brno 2008
- [Jir95] Jiroušek R.: *Metody reprezentace a zpracování znalostí v umělé inteligenci*. Vysoká škola ekonomická v Praze, 1995. 103s. ISBN 80-7079-701-0
- [MŠL93] Mařík V., Štěpánková O., Katanský J.: *Umělá inteligence 1*. 1. vyd. Praha : Academia, 2004. 264 s. ISBN 80-200-0496-3.
- [MŠL97] Mařík V., Štěpánková O., Katanský J.: *Umělá inteligence 2*. 1. vyd. Praha : Academia, 2003. 264 s. ISBN 80-200-0504-8.
- [PK99] Provozník I., Kozumplík J.: *Expertní systémy*, Brno 1999
- [RadW] Radovan.: *Expertní systémy – úvod* [www dokument] dostupný z: <http://radovan.blogger.cz/informatika/expertni-systemy/expertni-systemy---uvod>
- [Sed83] Sedláček V.: *Umělá inteligence -1. Úvod, metody řešení úloh, rezoluční metoda*. Univerzita J. E. Purkyně v Brně, 1987. 203 s.
- [ŠtýW] Štýbnarová L.: *Expertní systémy* [www dokument] dostupný z: http://ui.fpf.slu.cz/diplomky/znalostni_a_expertni_systemy/Welcome.html
- [Voj01] Vojta P.: *Expertní systém NPS 32*. Diplomová práce VUT FEI Brno, 2001

SEZNAM PŘÍLOH

Příloha 1 Výpis textové báze znalostí pravidl1.n32

Příloha 2 Výpis textové báze znalostí pravidl2.n32

Příloha 1

Výpis textové báze znalostí pravidl1.n32

.; Výběr vhodné grafické karty do PCIE. Pravidl1.n32 – jednoduchý zápis.

.ATI p=50%

.NVIDIA p=50%

.ZNACKA p=50% DK

;Jakou značku preferujete?

;*ATI [1]*

;*NVIDIA [2]*

ATI

NVIDIA

.AHRY p=50%

.AMEDIA p=50%

.AUCEL p=50% DK

;K čemu bude grafická karta primárně sloužit?

;*Hraní PC her [1]*

;*Multimédia (PC určené k přehrávání audia a videa)[2]*

AHRY

AMEDIA

< 51% NVIDIA

.NHRY p=50%

.NMEDIA p=50%

.NUCEL p=50% DK

;K čemu bude grafická karta primárně sloužit?

;*Hraní PC her [1]*

;*Multimédia (PC určené k přehrávání audia a videa)[2]*

NHRY

NMEDIA

< 51% ATI

.A3650S1 p=50% G

;ATI HD 3650 Sapphire

& 90% +ATI

& 90% -NVIDIA

& 90% -AHRY

& 90% +AMEDIA

.A3870X2P p=50% G

;ATI HD 3870 X2 Powercolor

& 90% +ATI

& 90% -NVIDIA

& 90% +AHRY

& 90% -AMEDIA

.N86GTI1 p=50% G

;NVIDIA 8600GT Inno3D

& 90% +NVIDIA

& 90% -ATI

& 90% -NHRY

& 90% +NMEDIA

.N98X2A2 p=50% G

;NVIDIA 9800X2 Asus

& 90% +NVIDIA

& 90% -ATI

& 90% +NHRY

& 90% -NMEDIA

#

Příloha 2

Výpis textové báze znalostí pravidl2.n32

.; Výběr vhodné grafické karty do PCIE. Pravidl2.n32 – rozšířený zápis.

.ATI p=50%

.NVIDIA p=50%

.ZNACKA p=50% DK

;Jakou značku preferujete?

;*ATI [1]*

;*NVIDIA [2]*

ATI

NVIDIA

.A700 p=50%

.A1300 p=50%

.A3300 p=50%

.A8000 p=50%

.ACENA p=50% DK

;Jakou má mít GK cenu?

;*700,- [1]*

;*1300,- [2]*

;*3300,- [3]*

;*8000,- [4]*

A700

A1300

A3300

A8000

< 51% NVIDIA

.N700 p=50%

.N1500 p=50%

.N3300 p=50%

.N11500 p=50%

.NCENA p=50% DK

;Jakou má mít GK cenu?

;*700,- [1]*

;*1500,- [2]*

;*3300,- [3]*

;*11500,- [4]*

N700

N1500

N3300

N11500

< 51% ATI

.A3450SP p=50% G

;ATI HD 3450 Sapphire

& 90% -NVIDIA

& 90% +A700

.A3650S1 p=50% G

;ATI HD 3650 Sapphire

& 90% -NVIDIA

& 90% +A1300

.A3870S1 p=50% G

;ATI HD 3870 Sapphire

& 90% -NVIDIA

& 90% +A3300

.A3870X2P p=50% G

;ATI HD 3870 X2 Powercolor
& 90% -NVIDIA
& 90% +A8000
.N84GSII p=50% G
;NVIDIA 8400GS Inno3D
& 90% -ATI
& 90% +N700
.N86GTII p=50% G
;NVIDIA 8600GT Inno3D
& 90% -ATI
& 90% +N1500
.N96GTII p=50% G
;NVIDIA 9600GT Inno3D
& 90% -ATI
& 90% +N3300
.N98X2A2 p=50% G
;NVIDIA 9800X2 Asus
& 90% -ATI
& 90% +N11500
#