

MASARYKOVA UNIVERSITA
FAKULTA INFORMATIKY



DIPLOMOVÁ PRÁCE

Expertní systém pro lékařskou diagnostiku
Jaroslav Měcháček

10. dubna 1996

Prohlašuji, že tuto diplomovou práci jsem napsal sám, s použitím uvedené literatury.

Děkuji Dr. Matyskovi a MUDr. Feitovi za poskytnuté informace, rady, připomínky, návrhy, literaturu a další věci potřebné k napsání této diplomové práce.

Obsah

1 Úvod	6
2 Neurčité znalosti a informace	7
3 Medicínské informace	9
4 Základní pojmy z teorie fuzzy množin	10
4.1 Základní operace s fuzzy množinami	10
4.2 Sémantika slov	11
4.3 Jazykové operátory	12
4.4 Jazyková proměnná	13
5 Expertní systémy	15
5.1 Tvorba expertního systému	16
5.2 Techniky používané v umělé inteligenci	16
5.2.1 Rámce	16
5.2.2 Systémy založené na pravidlech	17
6 Expertní systém pro lékařskou diagnostiku	17
6.1 Znalostní báze	18
6.1.1 Popis diagnostických znaků	19
6.1.2 Diagnostické jednotky	21
6.1.3 Slovní popis znaku	24
6.1.4 Snížení počtu diagnostických znaků	24
6.2 Grafické rozhraní	25
6.2.1 Diferenciální diagnostika	26
6.3 Inferenční mechanismus	26
6.3.1 Seřazení diagnóz	28
6.3.2 Rozhraní pro volání vyhodnocovacích metod	28
6.3.3 Použití významových funkcí	31
7 Závěr	31
A Definice významových funkcí pro vybrané znaky	34
B Ukázka grafického rozhraní	35

C DGM Help	36
C.1 DGM browser	36
C.2 Tvorba textu pro DGM browser	38

Seznam obrázků

1	Asociované funkce	14
2	Schématické znázornění jazykové proměnné teplota	19
3	Definice významových funkcí	20
4	Významové funkce	21
5	Popis diagnózy lichenoid actinic keratosis	23
6	Schéma vyhodnocování	27
7	Významové funkce $f_{teplota}$ a $f_{zvýšená\ teplota}$	32
8	DGM browser	37

Abstrakt

Diplomová práce se zabývá návrhem a implementací expertního systému pro lékařskou diagnostiku v logickém programovacím jazyce. Jde především o nalezení vhodné reprezentace lékařských znalostí a práci s neurčitou informací s využitím teorie fuzzy množin. Nemalá část práce byla věnována otázce komunikace s uživatelem. Práce vychází z již existujícího expertního systému pro diagnostiku jaterních biopsií.

1 Úvod

Cílem diplomové práce je vývoj lékařského konzultačního systému pro podporu diagnostiky se zaměřením na oblast patologie. Měl by být schopen pomoci při diagnostice onemocnění a na základě patologických nálezů poskytnout seznam diagnóz, které připadají v úvahu. Dále by měl poskytnout prostředky pro diferenciální diagnostiku.

Koncept expertního systému je založen na myšlence reprezentovat znalosti ve formě diagnostických jednotek, kde jedna diagnostická jednotka odpovídá popisu jedné diagnózy. Expertní systém potom srovnává popis patologických nálezů s popisy uloženými ve znalostní bázi, což odpovídá způsobu uvažování lékaře specialisty.

V medicíně se převážně setkáváme s nepřesnými a vágními informacemi, se kterými musí expertní systém umět zacházet. Teorie fuzzy množin poskytuje několik základních prostředků, jak tyto nepřesné informace popsat. Pomocí fuzzy teorie lze reprezentovat neohrazené medicínské entity jako fuzzy množiny, dále poskytuje vynikající nástroje pro reprezentaci vágních pojmů. Fuzzy logika poskytuje usuzovací metody schopné poskytovat přibližné závěry.

Jádro expertního systému je napsáno v logickém programovacím jazyce Prolog (Eclipse 3.3 nebo Sicstus 2.1), který poskytuje velmi dobré prostředky pro ukládání a zpracování znalostí. Grafické rozhraní je implementováno v jazyce tel/tk (Tel 7.4, Tk 4.0), který umožňuje poměrně rychlý vývoj grafické nadstavby a poskytuje standardní Motif vzhled a chování aplikace. Jádro i grafickou nadstavbu budeme v následujícím textu souhrnně označovat jako **DGM**.

2 Neurčité znalosti a informace

Žijeme ve světě, ve kterém musíme velmi často činit rozhodnutí a to na základě neurčitých a nepřesných znalostí a informací. Pokud nám má expertní systém při tomto rozhodování pomoci nebo dokonce rozhodovat za nás, je nutné, aby dokázal uchovávat a dále zpracovávat neurčité znalosti a na jejich základě se uměl rozhodnout.

Expertní systém používá svoje vlastní znalosti a dále informace o současné situaci, kterou řeší. Jak vlastní znalosti, tak informace o současné situaci mohou být nepřesné mnoha způsoby a expertní systém s nimi musí umět zacházet. Neurčitost se projevuje mnoha způsoby.

Pravděpodobnost

Pravděpodobnost může být chápána a formalizována mnoha způsoby. Pravděpodobnost jevu A můžeme chápat jako míru vyjadřující perspektivu nastání jevu A . Pravděpodobnost používáme v případě, kdy neznáme detailně všechny okolnosti, které mohou nastání daného jevu ovlivnit. Některé přístupy v umělé inteligenci se snaží zpracovávat neurčité informace výhradně na základě pravděpodobnosti. Otázkou je, jak získat čísla, vyjadřující pravděpodobnost zkoumaných jevů.

Můžeme říci, že čísla stanoví expert na základě svých zkušeností. Pro rozsáhlejší bázi znalostí je určení těchto čísel pro experta velmi těžké nebo zcela nemožné. Navíc nelze tento postup považovat za korektní.

Druhým přístupem je, že každé číslo určíme na základě systematického a mnohonásobného pozorování (měření) daného jevu, za neměnných podmínek (můžeme například statisticky vyhodnotit již existující pacientské záznamy). Zejména v lékařství, kde se objekty zkoumání (pacienti) i jejich chování podstatně liší, je asi nemožné zachovat neměnné podmínky pozorování. Dalším problémem je neúplnost pozorování, kdy některé jevy nemusí nutně nastat i když jejich pravděpodobnost není rovna nule. Například některé příznaky se při onemocnění vyskytují velmi vzácně, pokud se však objeví, jde určitě o dané onemocnění. V takových případech je potřeba zásah experta do znalostní báze, takže dojde ke kombinaci obou přístupů.

Nepřesnost

Přesnost existuje pouze díky abstrakci. Abstraktní modely reálného světa a jeho objektů (kružnice, bod) či předpokladů (ano, ne) jsou pouze umělé konstrukce. Reprezentují pouze ideální struktury a předpoklady, avšak na základě těchto abstrakcí je formována lidská mysl a znalosti jsou jejich výsledkem. Nekonečná přesnost je ve skutečném světě nemožná. Každé měření je tedy přesné jen do určité míry.

Neúplnost

Vlastnosti reálného světa se neustále mění a vyvíjí. Znalosti jsou tedy vždy nutně neúplné. Můžeme však najít spoustu dalších důvodů neúplnosti znalostí. Například získání některých informací je příliš drahé, nelze je získat v rozumně krátké době (viz. zkoumání genetického kódu), mohou být utajovány, současný stav přístrojového vybavení neumožňuje potřebnou přesnost měření nutného k získání těchto znalostí apod. Zejména v lékařství není možné některé výzkumy provádět z humaních důvodů.

Neúplnost a nepřesnost spolu do značné míry souvisí, tj. pokud je měření nepřesné, získáme neúplné znalosti.

Nekonzistence

Abstrakce vždy nemusí vést ke stejným výsledkům. Znalosti se mohou díky různé kultuře, náboženství, sociálních podmínek, zvyklostí a vzdělání značně lišit. To potom vede k tomu, že informace z různých zdrojů jsou nekonzistentní. Eliminovat nekonzistenci lze pouze v omezených systémech. Problém nekonzistence se v oblasti medicínských systémů vyskytuje velmi často. Zvyklosti a popis znalostí se velmi často a velmi značně liší v různých publikacích a je proto velmi obtížné je spojit dohromady.

Vágnost

Silným nástrojem lidského jazyka je schopnost používat vágní pojmy. Vágní pojem specifikuje určitou třídu objektů, ale hranice této třídy jsou velmi rozmazané. Nelze tedy pro každý prvek určit zda je, nebo není v dané třídě. Velmi silným prostředkem pro popis vágních pojmů je fuzzy teorie jejímž

autorem je L. A. Zadeh. Se základními pojmy fuzzy teorie se seznámíme v kapitole 4.

3 Medicínské informace

Informace, které má lékař k dispozici jsou velmi neurčité, nepřesné, neúplné a vágní. Míra neurčitosti je mnohem větší než v jiných oborech jako například fyzika. Navzdory tomu je lékař většinou schopen určit správnou diagnózu. Pacientská data si můžeme rozdělit do následujících kategorií:

- *Informace sdělené pacientem:* Jsou velmi subjektivní a mohou obsahovat symptomy vsugerované nebo simulované. Pacient může zamlčet důležité informace o prodělaných nemocech nebo operacích, které mohou být velmi důležité pro správné určení onemocnění.
- *Fyzické vyšetření:* Zde může lékař získat více méně objektivní data. Bohužel i lékaři dělají někdy chyby a mohou přehlédnout důležité příznaky. Hranice mezi normálním a patologickým nálezem nejsou přesné.
- *Laboratorní výsledky:* Laboratorní výsledky lze považovat za objektivní data. Avšak i zde se mohou objevit chyby v měření, organizační problémy (chybné označení vzorků), chybné chování pacienta před vyšetřením, což může vést ke zcela nekorektním datům. Navíc hranice mezi normálním a patologickým výsledkem nejsou vždy přesně vymezené. Jsou zde vždy hraniční hodnoty, které nemohou být označeny ani jako normální, ani jako patologické.
- *Rentgenové, ultrazvukové a jiné výsledky:* Tyto výsledky znovu závisí na správné interpretaci lékaře.

Diagnóza má dvě hlavní složky:

klasifikace příslušné onemocnění pojmenuje. Při klasifikaci musí lékař pracovat s velkým počtem kvalitativních i kvantitativních znaků a mít přehled o velkém množství onemocnění.

gradace určuje stupeň onemocnění. Při gradaci lékař uvažuje pouze malý počet kvantitativních znaků. Lékař zde nepotřebuje tak rozsáhlé znalosti jako při klasifikaci, spíše je potřeba cvik a určitý cit pro určení stupně onemocnění.

Proto bude expertní systém přínosnější pro experta spíše při klasifikaci, kdy může během krátké doby provést srovnání se stovkami diagnóz popsaných ve znalostní bázi.

4 Základní pojmy z teorie fuzzy množin

Základním pojmem fuzzy teorie je pojem fuzzy množina. Fuzzy množina je třídou objektů, jež nemá ostře vymezené hranice, které by jasně určovaly, které objekty do ní patří a které ne. Namísto toho je každému objektu přiřazen stupeň příslušnosti, který může například vyjadřovat míru našeho přesvědčení, že daný prvek patří do této třídy.

Nechť U je množina a $\mathcal{L} = \langle L, \vee, \wedge, 1, 0 \rangle$ je svaz. Pak fuzzy množina A v univerzu U je funkce

$$A : U \rightarrow L$$

V následujícím textu budeme předpokládat, že $L = \langle 0, 1 \rangle$. Funkce A přiřazuje každému prvku $u \in U$ číslo $Au \in \langle 0, 1 \rangle$, které nazýváme *stupeň příslušnosti* prvku u do fuzzy množiny A . Nabývá-li funkce A pouze hodnot 0,1, stává se “klasickou” množinou. Explicitně budeme fuzzy množinu zapisovat následovně:

$$A = \bigcup_{u \in U} Au / u$$

Je-li A fuzzy množina v univerzu U , pak píšeme

$$A \check{C} U$$

A je podmnožinou B (nebo A je menší nebo rovna B) právě tehdy když pro $\forall u \in U$ platí $Au \leq Bu$

4.1 Základní operace s fuzzy množinami

Nechť $A \check{C} U$, $B \check{C} U$ jsou fuzzy množiny. Pak definujeme základní operace s fuzzy množinami takto:

sjednocení

$$A \cup B = \bigcup_{u \in U} (Au \vee Bu) / u \quad (1)$$

průnik

$$A \cap B = \bigcup_{u \in U} (Au \wedge Bu) / u \quad (2)$$

doplňk

$$\bar{A} = \bigcup_{u \in U} (1 - Au) / u \quad (3)$$

součin

$$AB = \bigcup_{u \in U} (AuBu) / u \quad (4)$$

4.2 Sémantika slov

Sémantika přirozeného jazyka je více či méně vágní. Proto se jeví vcelku výhodné modelovat sémantiku (objektových) slov jazyka pomocí fuzzy množin. Význam slov budeme ztotožňovat s vymezenou třídou prvků univerza U , obsahující generické prvky, které mají vztah k danému kontextu. Množinu K generických prvků nazýváme *jaderným prostorem*.

Dále $\mathcal{F}_{\mathcal{L}}(K) = \{A : \forall A \tilde{C} K\}$ a E_K je nejmenší množina obsahující K a uzavřená vzhledem ke kartézskému součinu a operátoru $\mathcal{F}_{\mathcal{L}}$.

Prvky univerza E_K budeme ztotožňovat s významem slov daného jazyka. Nechť $U \subseteq E_K$, \mathcal{T} je množina slov, popř. slovních spojení nějakého jazyka \mathbf{L} . Jazyku \mathbf{L} přiřadíme sémantickou fuzzy relaci

$$S \tilde{C} \mathcal{T} \times U \quad (5)$$

kde $S\langle \mathcal{A}, u \rangle \in \langle 0, 1 \rangle$, $\mathcal{A} \in \mathcal{T}$, $u \in U$.

Nechť S je relace (5) $\mathcal{A} \in \mathcal{T}$ je slovo. *Význam slova* \mathcal{A} definujeme jako fuzzy množinu

$$M(\mathcal{A}) = \bigcup_{u \in U} S\langle \mathcal{A}, u \rangle / u \quad (6)$$

Deskripční množinu $D(u)$ prvku $u \in U$ definujeme jako fuzzy množinu

$$D(u) = \bigcup_{\mathcal{A} \in \mathcal{T}} S\langle \mathcal{A}, u \rangle / \mathcal{A} \quad (7)$$

Slova $\mathcal{A}_1, \mathcal{A}_2 \in \mathcal{T}$ jsou synonyma, jestliže $M(\mathcal{A}_1) = M(\mathcal{A}_2)$.

4.3 Jazykové operátory

Jazykovými operátory (tj. příslovci velmi, mírně, více méně ...) můžeme podrobněji specifikovat to, co pozorujeme, popisujeme nebo chceme říct. Každému jazykovému operátoru m přiřadíme asociovanou funkci

$$Asf_m : \langle 0, 1 \rangle \rightarrow \langle 0, 1 \rangle$$

Pokud $\mathcal{A} \in \mathcal{T}$ je slovo, $m \in \mathcal{T}$ jazykový operátor, $M(\mathcal{A}) = \bigcup_{u \in U} S\langle \mathcal{A}, u \rangle / u$ je význam slova \mathcal{A} , pak význam výrazu $m\mathcal{A}$ definujeme jako

$$M(m\mathcal{A}) = \bigcup_{u \in U} Af_{s_m}(S\langle \mathcal{A}, u \rangle) / u$$

Nejdříve si definujeme pomocné funkce potřebné k definici asociovaných funkcí.

koncentrace

$$CON(x) = x^2 \tag{8}$$

dilatace

$$DIL(x) = -CON(\neg x) = 2x - x^2 \tag{9}$$

nebo

$$DIL(x) = x^{0.5} \tag{10}$$

intenzifikace

$$INT(x) = \begin{cases} 2x^2 & \text{pro } x \in \langle 0, 0.5 \rangle \\ 1 - 2(1 - x)^2 & \text{pro } x \in \langle 0.5, 1 \rangle \end{cases} \tag{11}$$

normalizace

$$NORM(f(x)) = \frac{f(x)}{\hat{x}} \tag{12}$$

kde $\hat{x} = \bigvee_{x \in \langle 0, 1 \rangle} f(x) \neq 0$, $f : L \rightarrow L$.

Koncentrace redukuje stupně příslušnosti, přičemž redukce je relativně malá pro velká Au a značně velká pro malé Au . Dilatace má opačný účinek než koncentrace. Intenzifikace zvyšuje hodnoty $Au > 0.5$ a snižuje hodnoty $Au < 0.5$. Tato operace tedy snižuje neostrost (rozmazanost) fuzzy množiny.

Nyní již můžeme přistoupit k definici asociovaných funkcí jazykových operátorů:

velmi (very):

$$Afs_{velmi}(x) = CON(x) \quad (13)$$

Toto je v aplikacích fuzzy množin nejpoužívanější operátor. Hodnoty funkce příslušnosti se snižují, kromě $x = 1$, která zůstává stejná. V tom je určitá neadekvátnost, neboť kromě koncentrace kolem 1 by se měla funkce příslušnosti mírně posunout do prava. Je-li např. člověk vysoký 180 cm “velký” se stupněm příslušnosti 1, pak není zcela zřejmé, že je “velmi velký” také se stupněm 1. Navíc by se tento posun měl s opakováním “velmi” zmenšovat, jinak by po dostatečném počtu opakování “velmi” výraz “velmi . . . velmi velký” znamenal, že se stupněm 1 by musel mít člověk zhruba 6 m.

více méně (more or less):

$$Afs_{víceméně}(x) = DIL(x) \quad (14)$$

značně (highly):

$$Afs_{značně}(x) = x^3 \quad (15)$$

zhruba (roughly):

$$Afs_{zhruba}(x) = DIL(DIL(x)) \quad (16)$$

spíše (rather):

$$Afs_{spíše}(x) = INT(CON(x)) \quad (17)$$

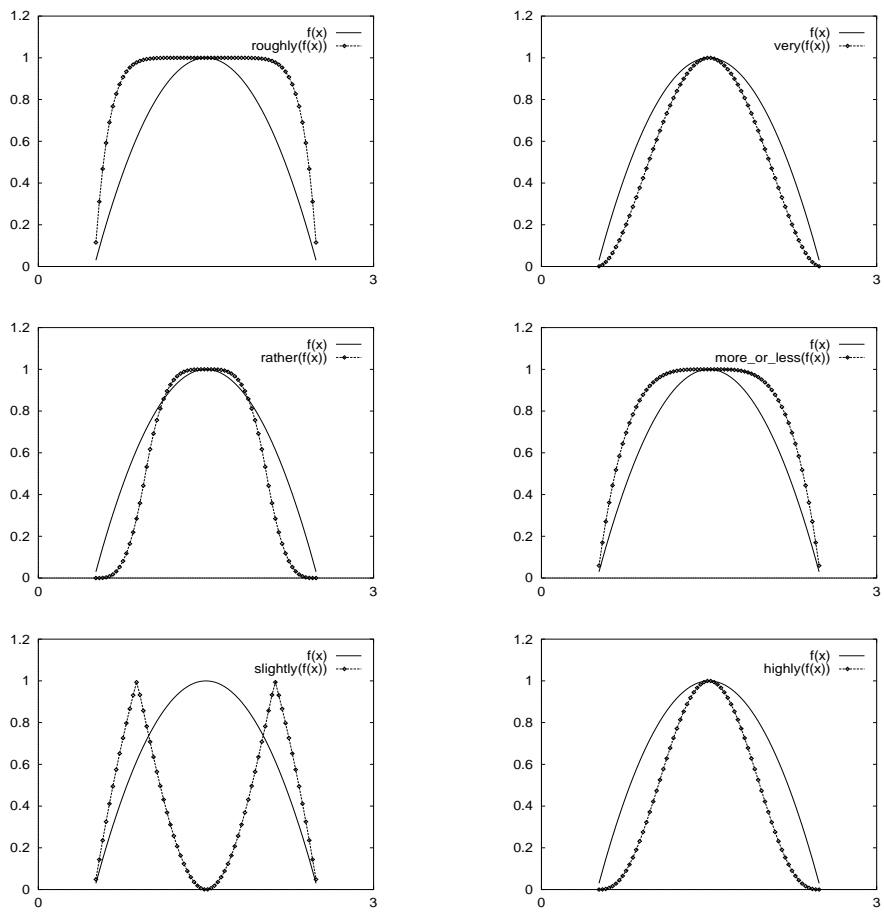
mírně (slightly):

$$Afs_{mírně}(x) = NORM(x \wedge \neg CON(x)) \quad (18)$$

Graficky jsou asociované funkce znázorněny na obrázku 1.

4.4 Jazyková proměnná

je proměnná, jejíž hodnoty jsou slova a významy slov jsou definovány pomocí fuzzy množin. Používání slov jako hodnot proměnných místo čísel je motivováno tím, že jazykový popis skutečnosti bývá zpravidla méně konkrétní než popis pomocí čísel. Například věta “Petr je mladý” je méně konkrétní než



Obrázek 1: Asociované funkce

věta “Petrovi je 25 let”. Zde slovo mladý můžeme chápat jako hodnotu jazykové proměnné věk. Jazyková hodnota proměnné je přitom stejně důležitá jako třeba hodnota numerická. Hodnoty jazykové proměnné budeme nazývat *termy*.

Obyčejná proměnná je charakterizována trojicí $\langle X, U, R(X) \rangle$, kde X je jméno proměnné, U je universum, $R(X) \subseteq U$ je obor hodnot proměnné, neboli ohraničení proměnné X .

Je-li ohraničení $R(X)$ proměnné X fuzzy množinou v U , $R(X) \tilde{\subset} U$, nazveme X *fuzzy proměnnou*. Potom prvek $x \in U$ je hodnotou proměnné X ve stupni $R(X)x$.

Jazyková proměnná je charakterizována pěticí $\langle \mathcal{K}, \mathcal{T}(\mathcal{K}), U, G, \mathcal{M} \rangle$, kde \mathcal{K} je jméno proměnné

$\mathcal{T}(\mathcal{K})$ je množina termů, tj. jazykových hodnot proměnné \mathcal{K}

U je universum

G je syntaktické pravidlo, které generuje prvky $\mathcal{A} \in \mathcal{T}(\mathcal{K})$

\mathcal{M} přiřazuje každému termu $\mathcal{A} \in \mathcal{T}(\mathcal{K})$ jeho význam $\mathcal{M}(\mathcal{A}) \tilde{\subset} U$.

Podle definice tedy konkrétní hodnota proměnné X generovaná pravidlem G je term. Term, který se skládá z jednoho slova nebo několika slov, které se vyskytují vždy spolu budeme nazývat *atomický term*.

Pro naše účely si můžeme definici jazykové proměnné zjednodušit jako pěticí $\langle \mathcal{K}, \mathcal{T}(\mathcal{K}), \mathcal{T}, U, \mathcal{M} \rangle$, kde

\mathcal{T} obsahuje pouze atomické termy.

\mathcal{M} přiřazuje každému atomickému termu jeho význam.

$\mathcal{T}(\mathcal{K})$ obsahuje atomické termy $\mathcal{A} \in \mathcal{T}$ a výrazy tvaru $m\mathcal{A}$, kde m je jazykový operátor.

Jazyková fuzzy proměnná je šestice $\langle \mathcal{K}, \mathcal{T}, \mathcal{T}(\mathcal{K}), U, \mathcal{M}, R(\mathcal{K}) \rangle$, kde

$R(\mathcal{K}) \tilde{\subset} \mathcal{T}(\mathcal{K})$.

5 Expertní systémy

Expertní systém je počítačový program, který napodobuje chování a uvažování experta ve specifické oblasti (doméně). Základní charakteristiky expertního systému jsou:

1. schopnost podporovat experta
2. reprezentace specifických informací a jejich zpracování způsobem jakým uvažuje expert
3. schopnost zpracovávat neúplné a nepřesné znalosti

Expertní systém můžeme rozdělit na tři základní části:

- rozhraní pro komunikaci s uživatelem, umožňující klást otázky a získávat odpovědi.
- znalostní báze obsahuje soubor faktů a pravidel týkajících se určité domény.
- vyvozovací mechanismus (inference engine), který generuje odpovědi.

Protože je znalostní báze oddělena od zbytku systému, může být dále rozšiřována a stejný systém může být použit v různých oblastech nahrazením jedné znalostní báze za druhou.

5.1 Tvorba expertního systému

Prvním krokem při tvorbě expertního systému je zvolit “správný problém”, definovat cíl, kterého chceme dosáhnout, identifikovat zdroj znalostí. Problém by měl mít dobře ohraničenou doménu, kterou chceme zpracovat, abychom nebyli zaplaveni velkým množstvím alternativ.

Dalším krokem je získání znalostí od experta, zvolení jejich vhodné reprezentace a vytvoření znalostní báze. Zvolení reprezentace je závislé především na typu znalostí. Může to být například reprezentace pomocí rámců, pravidel, sítě a podobně. Získávání znalostí je iterativní proces, který zahrnuje mnoho konzultací s expertem tak, aby většina znalostí byla zahrnuta ve znalostní bázi.

Potom následuje návrh a implementace inferenčního mechanismu, jeho testování a opravy znalostní báze. Testování můžeme provést tak, že necháme expertní systém řešit problémy, jejichž řešení předem známe, nebo ho můžeme poskytnout k používání ostatním expertům, kteří se zabývají danou oblastí.

5.2 Techniky používané v umělé inteligenci

5.2.1 Rámce

Rámce umožňují popis objektů a sdílení těchto popisů mezi objekty. To umožňuje zachování integrity systému a zabraňuje vzniku redundance dat.

Rámec je datová struktura, která může být použita k popisu atributů, omezení a typu objektu nebo třídy objektů.

- Atributy rámce (někdy též nazývané sloty) slouží k ukládání hodnot, popisujících vlastnosti daného objektu
- Odkazy (links) definují vztahy mezi rámci (vztah mezi třídou a podtřídou)
- Facety specifikují vlastnosti atributů. Facet může představovat omezení atributu, datový typ atributu, komentáře, čas přidání.

5.2.2 Systémy založené na pravidlech

sestávají z množiny pravidel tvaru *if X then Y*. Levá strana pravidla (LHS) popisuje situaci, pravá strana pravidla (RHS) popisuje

- akci, která se má provést — operativní pravidlo
- závěr, který vyplývá ze splněných předpokladů pro LHS — logické pravidlo

V systémech založených na pravidlech rozeznáváme dvě základní strategie

- Dopředné řetězení: inferenční stroj vybírá pravidla jejichž LHS uspěje se současným stavem paměti. Pokud je vybráno pouze jedno pravidlo, provede se jeho RHS. Pokud je vybráno více pravidel, je třeba z této konfliktní množiny vybrat pouze jedno pravidlo. Tento výběr obvykle bývá založen na některé z následujících metod nebo jejich kombinaci
 - je provedeno pravidlo, které bylo vybráno jako první
 - upřednostnění pravidla, které bylo uloženo do paměti jako poslední
 - upřednostnění konkrétnějších faktů
- Zpětné řetězení: hypotéza LHS je splněna, pokud jsou splněny podcíle obsažené v RHS

6 Expertní systém pro lékařskou diagnostiku

DGM je implementován pod operačním systémem UNIX, s grafickou nadstavbou pro X Window system. Přenositelnost mezi jednotlivými verzemi UNIXu je závislá pouze na dostupnosti Eclipse (nebo Sicstusu) na dané verzi. Pro popis znalostí a komunikaci s uživatelem byl použit anglický jazyk.

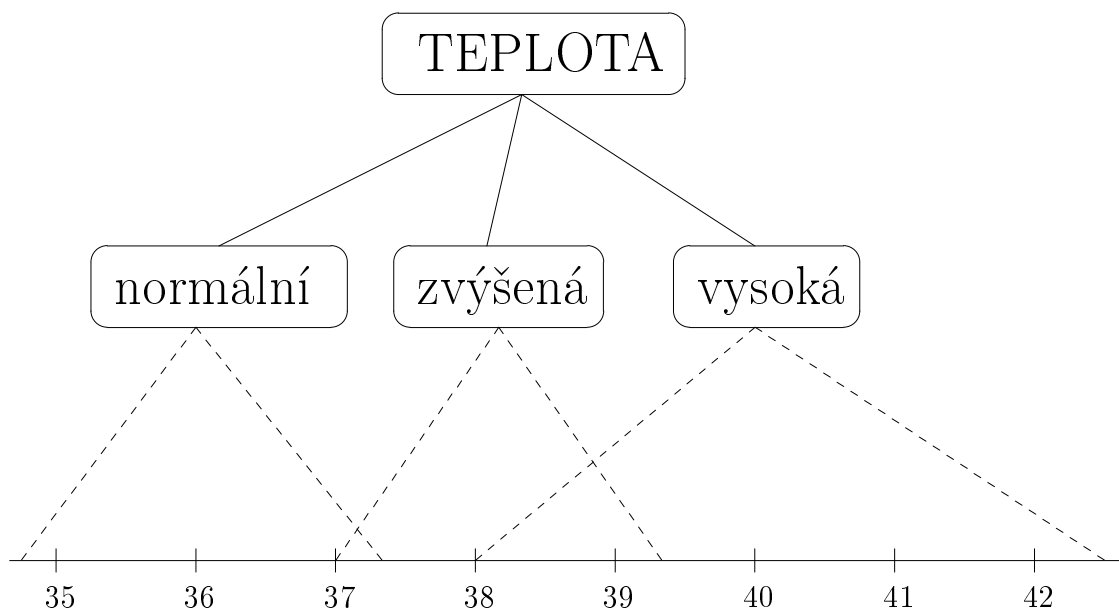
6.1 Znalostní báze

Základem každého expertního systému je znalostní báze. Pokud je báze znalostí chybná nebo neúplná, tak také výsledky expertního systému budou chybné nebo neúplné, nezávisle na tom jak kvalitní inferenční mechanismus je implementován.

Ačkoliv je většina dnešních expertních systémů založena na pravidlech, expertní systém prezentovaný v této diplomové práci staví především na diagnostických jednotkách popisujících jednotlivé diagnózy. Pokusím se shrnout důvody, které vedly k tomuto řešení (viz. [3]).

- Nevýhody systémů založených na pravidlech
 - Medicínské informace jsou těžko reprezentovatelné (pouze) pomocí pravidel.
 - Informace pro naplnění báze znalostí jsou získávány především z odborné medicínské literatury. Tyto informace jsou převážně charakteru popisného, málokdy lze najít informace ve tvaru pravidel.
 - Je velmi těžké zajistit konzistenci pravidel, zejména v systémech kde jsou neustále přidávány nové informace.
- Expertní systém se snaží napodobit uvažování lékaře specialisty. Pokud se podíváme na to, jak uvažuje lékař (patolog), zjistíme, že “nezpracovává” spoustu pravidel, ale spíše se snaží ve své paměti najít obraz diagnózy, která by odpovídala pozorovanému objektu.
- Informace uložené ve tvaru diagnostických jednotek mohou být využity k difereciální diagnostice.

Nyní popíšeme, jak jsou medicínská data reprezentována v systému DGM pomocí logického programovacího jazyka Prolog. Znalostní báze, obsahující informace z jisté lékařské *oblasti*, se skládá z diagnostických jednotek. Dalším důležitým pojmem je *diagnostický znak*. Pro každou oblast definujeme množinu diagnostických znaků, které k ní mají vztah. Např. pro oblast jaterních biopsií má smysl definovat znaky barva, fragmentace, tvar, naopak dušnost, nebo sedimentace jsou znaky irelevantní a nemohou k diagnostikování onemocnění nijak přispět. Diagnostické znaky jsou dále sdruženy do logických celků nazývaných topografie.



Obrázek 2: Schématické znázornění jazykové proměnné teplota

6.1.1 Popis diagnostických znaků

Nejdříve je třeba systému sdělit, že znak daného jména je relevantní pro danou oblast:

```
sign_def(jméno, typ, seznam_hodnot).
```

Diagnostický znak můžeme ztotožnit s jazykovou proměnnou. Potom

jméno je jméno jazykové proměnné

seznam_hodnot je seznam atomických termů.

typ udává, zda je možno atomické termy uspořádat. Pokud například seznam hodnot je [muž, žena, dítě] nemá smysl hovořit o uspořádání. Naopak [malý, velký, obrovský] je již možno uspořádat.

Jako příklad nám poslouží jazyková proměnná s názvem **teplota** (myšleno tělesná teplota). Potom atomické termy budou **normální** (normal), **zvýšená** (increased) a **vysoká** (high). Jaderný prostor bude interval reálných čísel $\langle 34, 42 \rangle$. Schématicky je jazyková proměnná teplota znázorněna na obrázku 2. Definice diagnostického znaku teplota bude vypadat následovně.

```

:-mean_def(teplota,q:s:[muž,žena],
  [ normální:[(35,1),(36,1),(37.4,0)],
    zvýšená:[(37,0),(38,1),(38.3,1),(39.0)],
    vysoká:[(38,0),(39.3,1),(42,1)]
  ]
).
:-mean_def(teplota,q:s:[dítě],
  [ normální:[(35,1),(36,1),(37.4,0)],
    zvýšená:[(37,0),(38,1),(38.3,1),(38.8,0)],
    vysoká:[(37.7,0),(39.3,1),(42,1)]
  ]
).

```

Obrázek 3: Definice významových funkcí

```
sign_def(teplota,y,[normální,zvýšená,vysoká]).
```

Zbývá nám ještě definovat význam atomických termů. Jako univerzum U použijeme množinu reálných čísel R . Každému atomickému termu O přiřadíme významovou funkci $f_O : R \rightarrow \langle 0, 1 \rangle$ a význam $\mathcal{M}(O)$ termu O definujeme jako

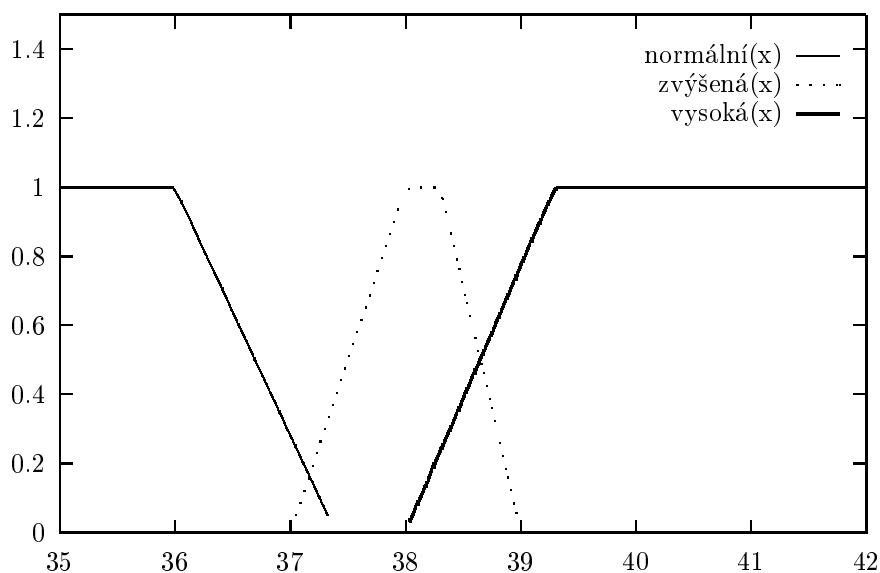
$$\mathcal{M}(O) = \bigcup_{u \in R} f_O(u) / u$$

DGM umožňuje zadat významové funkce aproximací pomocí úseček, a to tak, že jsou zadány body ležící na křivce. Body musejí být zadány vzestupně, uspořádané podle x-ové souřadnice. Předpokládá se, že průběh křivek může být závislý na jiném diagnostickém znaku. Například mnoho hodnot je třeba odlišně interpretovat pokud se jedná o dospělého jedince a jinak pokud jde o dítě. Na obr. 3 je definován diagnostický znak **teplota**, a významové funkce pro muže, ženu a dítě.

Funkce $f_{normální}$, $f_{zvýšená}$, $f_{vysoká}$ jsou znázorněny na obrázku 4. Přiřazení významů jednotlivým atomickým termům není povinné a někdy není ani smysluplné.

Další věc, kterou je nutno ve znalostní bázi specifikovat je vztah mezi diagnostickým znakem a danou diagnózou. Nechť J je jméno znaku a D je jméno diagnózy. Tento vztah je vyjádřen ve tvaru

J : popis - modifikátor : důležitost



Obrázek 4: Významové funkce

popis je nejdůležitější část. Vyjadřuje (většinou slovním popisem), jakých hodnot znak nabývá nebo naopak nenabývá, pokud se skutečně jedná o diagnózu D. Slovní popis má pevně danou syntaxi, kterou se budeme zabývat později. Může vypadat například takto:

- 'is always very high'
- 'is never none'
- 'is usually increased or high'

modifikátor může více upřesnit popis.

důležitost určuje, jak je tento znak důležitý pro určení, zda se jedná o diagnózu D. Nejdůležitější jsou znaky diagnostické, tj. znaky, které pokud nabudou určitých hodnot, potom se s vysokou jistotou jedná o diagnózu D (nebo je naopak vyloučena).

6.1.2 Diagnostické jednotky

Nyní, kdy už umíme popsat jednotlivé znaky a jejich vztah k diagnózám, můžeme popsat danou diagnózu. Každá diagnostická jednotka odpovídá jedné diagnóze a je popsána prologovským termem následujícího tvaru:

```

dg(Jméno_diagnózy <-- Dg_od_které_dědím_nedefinované_topo_celky, [
  Jméno1_topo_celku <-- Jméno_dg_od_které_dědím_tento_topo_celek :
  [
    Jméno1_dg_znaku <-- Jméno_dg_od_které_dědím_tento_znak
      : popis_znaku - modifikátor : důležitost_znaku,
    Jméno2_dg_znaku <-- Jméno_dg_od_které_dědím_tento_znak
      : popis_znaku - modifikátor : důležitost_znaku,
    .....
  ],
  Jméno2_topo_celku <-- Jméno_dg_od_které_dědím_tento_topo_celek :
  [ ..... ]),
  .....
]).

```

Popisy znaků (a jejich důležitost) mohou být děděny na úrovni:

- jednotlivých znaků – dědíme popis znaku od jiné diagnózy
- jednotlivých topografických celků – dědíme popis celého topografického celku od jiné diagnózy
- diagnostické jednotky – zdědíme popis celé diagnózy

Na obrázku 5 je znázorněn příklad popisu diagnózy lichenoid actinic keratosis. Jde o kožní onemocnění. Topografické celky zde představují jednotlivé vrstvy pokožky (stratum corneum, stratum spinosum, junction, papillary dermis a upper dermis). V rámci každého topografického celku je popsán vzhled jednotlivých znaků. norma je diagnostická jednotka popisující normální kůži. Diagnostickou jednotku můžeme číst následovně:

- Topografické celky, které nejsou uvedeny, odpovídají popisu normální kůže.
- Znaky, které nebyly popsány v rámci topografického celku junction jsou zděděny od popisu diagnózy lichen simplex chronicus.
- Znak lymphocytes z topografického celku papillary dermis je taktéž zděděn od popisu diagnózy lichen simplex chronicus.
- Nepopsané znaky v topografiích stratum corneum, stratum spinosum, papillary dermis a upper dermis jsou zděděny od popisu normální kůže.

```
:-dg( lichenoid_actinic_keratosis <-- norma, [
  stratum_corneum : [
    keratotic_layer : 'is normokeratotic',
    parakeratosis : 'is higher than none'
  ],
  stratum_spinosum <-- norma : [
    acanthosis : 'is irregular',
    acanthosis_grade : 'is slight or moderate',
    polarity : 'is lost',
    keratinocytic_atypia : 'is moderate'
  ],
  junction <-- lichen_simplex_chronicus : [
    infiltrate_density : 'is dense',
    lymphocytes : 'dominate',
    plasmocytes : 'are admixed'
  ],
  papillary_dermis <-- norma : [
    infiltrate_density : 'is dense',
    lymphocytes <-- lichen_simplex_chronicus,
    plasmocytes : 'are admixed'
  ],
  upper_dermis <-- norma : [
    infiltrate_density : 'is dense',
    infiltrate_distribution : 'is bandlike',
    lymphocytes : 'dominate',
    plasmocytes : 'are admixed',
    solar_elastosis : 'is moderate or heavy'
  ]
]).
```

Obrázek 5: Popis diagnózy lichenoid actinic keratosis

6.1.3 Slovní popis znaku

Nyní se vrátíme k slovnímu popisu diagnostických znaků. Syntaxe je poměrně jednoduchá,

(sloveso) (příslovce častosti) ostatní členy

avšak jeví se jako zcela dostatečná a zatím není důvod k jejímu obohacení. Slovní popis může obsahovat:

- atomické termy, uvedené v definici znaku
- slovesa *is, are*
- příslovce častosti *always, usually, often, sometimes, unusually, never*
- slovní spojení *higher than, lower than*
- Jazykové operátory *highly, roughly, rather, slightly, very*
- logickou spojku *or*

6.1.4 Snížení počtu diagnostických znaků

V rámci dané lékařské domény (například onemocnění kůže) lze někdy identifikovat (obvykle kvalitativní) diagnostické znaky, pomocí kterých lze určitou skupinu onemocnění vyloučit. DGM umožňuje tyto znaky označit a určit, které diagnózy lze vyloučit, pokud znak nabude jisté hodnoty. Pojem vyloučení zde neslouží k tomu, aby DGM takto označené diagnózy vyřadil z dalšího zpracování, protože uživatel může zadat chybnou hodnotu, může se jednat o ojedinělý průběh onemocnění apod. DGM pouze na základě této znalosti může během komunikace s uživatelem označit některé znaky jako méně důležité (důležitostí znaku zde chápeme to, jak přínosné bude zadání hodnoty znaku pro další rozlišení diagnóz). Pro lepší pochopení uvedeme příklad:

```
: -hide_dg(stratum_corneum:puchýř, [ne], [dg1, dg2, dg3]).
```

Predikát **hide_dg/3** určuje, že pokud uživatel přiřadí znaku *puchýř* v topografii *stratum_corneum* hodnotu *ne*, potom DGM může vyloučit puchýřnaté diagnózy *dg1, dg2, dg3*. To se projeví tak, že DGM vezme zbylé diagnózy a znaky, které nevedou k rozlišení mezi těmito diagnózami jsou označeny jako méně významné (používá se barevného odlišení znaků). Opticky se tedy zmenší počet znaků, kterým by měl uživatel přiřadit nějakou hodnotu. Druhý

argument může obsahovat více hodnot, při kterých budou diagnózy, uvedené ve třetím argumentu, vyloučeny.

V případě, že uživatel sdělil expertnímu systému, že puchýře nejsou, jeví se jako rozumné nezobrazovat vůbec znaky, které přímo s puchýři souvisí. Uvedeme další příklad:

```
: -hide_sign(stratum_corneum:puchýř, [ne],  
  [stratum_corneum:velikost_puchýře,  
  stratum_corneum:barva_puchýře]).
```

Tímto způsobem expertnímu systému sdělíme, aby nezobrazoval diagnostické znaky *velikost_puchýře* a *barva_puchýře* pokud puchýře nejsou.

6.2 Grafické rozhraní

Veškerá komunikace uživatele s expertním systémem se děje prostřednictvím grafické nadstavby v X Window systému. Rozhraní je rozděleno do dvou oken. Základní okno slouží k manipulaci se znalostní bází a k popisu jednotlivých znaků podle subjektivního dojmu uživatele, výsledků laboratorních měření apod. Výsledky získané inferenčním mechanismem jsou zobrazovány v okně druhém.

Na rozdíl od systémů založených na pravidlech, kde bývá zvykem, že expertní systém klade uživateli otázky, DGM ponechává rozhodnutí o tom jaké informace budou zadány, pouze na uvážení uživatele. V některých případech je DGM schopen průběžně na základě vstupů od uživatele označit některé znaky jako méně významné. Popis takto označených znaků není příliš přínosný a uživatel se může soustředit na znaky takto neoznačené. Například pokud uživatel zadá, že puchýře nejsou, potom DGM označí jako nevýznamné ty znaky, které nevedou k rozlišení mezi nepuchýřnatými chorobami.

Práce uživatele se systémem vypadá obvykle následovně: Nejdříve si uživatel zvolí znalostní bázi, se kterou chce pracovat (například znalosti z oblasti onemocnění kůže). DGM uloží tuto bázi do operační paměti, vytvoří seznam znaků a topografií, vyskytujících se ve znalostní bázi, a zobrazí je v hlavním okně. V horní části jsou zobrazeny topografické celky, v dolní části diagnostické znaky, patřící do topografie zvolené uživatelem.

Nyní si uživatel zvolí topografický celek a začne přiřazovat diagnostickým znakům jejich hodnoty. Znaků může být přiřazena hodnota:

- slovní (např. teplota je zvýšená)

- číselná (např. teplota je 38.5)
- označit znak jako normální
- označit znak jako nedefinovaný

Hodnoty mohou být zadávány hromadně, např. lze označit skupinu znaků jako normální nebo nedefinované.

Každému znaku, navíc může přiřadit slovní spojení, které vyjadřuje uživatelskou jistotu (možná, určitě), nebo jazykový operátor (velmi, zhruba, mírně).

V průběhu zadávání je uživateli k dispozici nápověda. Stiskem tlačítka help si může kdykoli vyvolat informace o znaku, topografii nebo celé tematické oblasti. Systém nápovědy je podrobněji popsán v příloze C.

Pokud uživatel zadá všechny hodnoty, které považuje za důležité nebo patologické, může spustit inferenční stroj. Ten se na základě báze znalostí a zadaných hodnot snaží vybrat diagnózy, které nejvíce odpovídají popisu a poskytne jejich seznam uživateli spolu s dalšími informacemi jako je skóre, které ta která diagnóza získala.

Pokud není uživatel s výsledky spokojen, může celý postup opakovat (tj. zadat další hodnoty nebo je pozměnit). Další možností je použít modul pro diferenciální diagnostiku, který může být velmi užitečný v případech, kdy se uživatel potřebuje rozhodnout mezi několika diagnózami které připadají v úvahu.

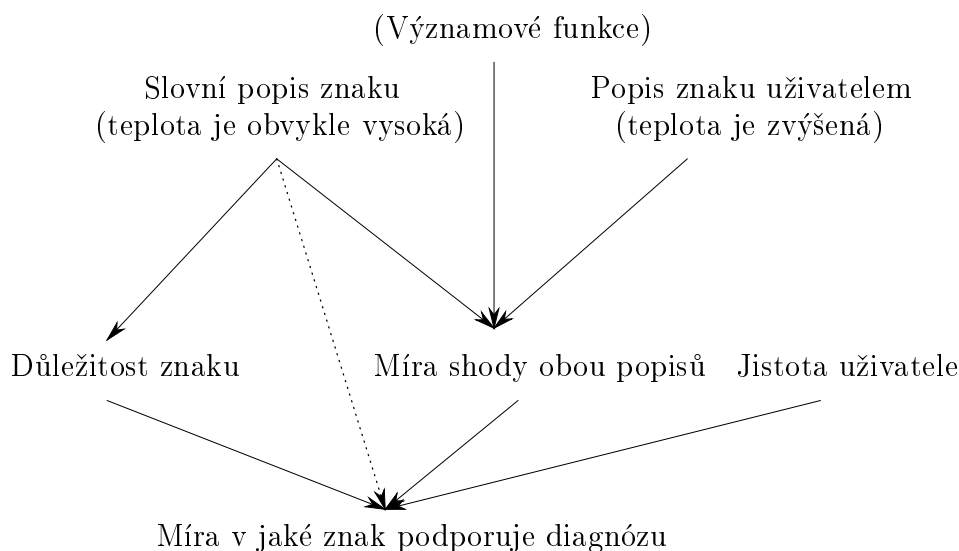
6.2.1 Diferenciální diagnostika

Někdy může dojít k situaci, kdy inferenční stroj přiřadí na základě informací od uživatele dvěma nebo více diagnózám téměř identické skóre. Pokud se uživatel nechce spokojit s takovým výsledkem, má k dispozici modul pro diferenciální diagnostiku. Zde může označit dvě nebo více diagnóz, se kterými chce pracovat. Systém projde diagnostické jednotky, popisující vybrané diagnózy a zobrazí v hlavním okně pouze znaky, v jejichž popise se odlišují.

Uživatel má nyní k dispozici omezenou množinu znaků a nemusí se zabývat ostatními znaky, které by stejně nevedly k dalšímu rozlišení diagnóz.

6.3 Inferenční mechanismus

Cílem při vyhodnocování informací získaných od uživatele je vybrat z velkého množství diagnóz malou množinu takových diagnóz, které na základě popisu připadají v úvahu. Tento problém můžeme převést na uspořádání diagnóz.



Obrázek 6: Schéma vyhodnocování

Výsledkem je seřazený seznam diagnóz, kde na začátku jsou diagnózy, jejichž popis nejvíce odpovídá popisu uživatele. Vyhodnocení můžeme rozdělit na dvě části:

- Nezávisle pro každý znak a diagnózu určíme, zda znak diagnózu podporuje, nebo zda je proti a dále míru této podpory.
- Uspořádání diagnóz na základě výsledků z části první.

Nejdříve se budeme zabývat částí první. Vezměme tedy diagnózu D a diagnostický znak Z . Nyní je třeba určit, zda znak Z podporuje diagnózu D a v jaké míře. Míra podpory diagnózy je reálné číslo z intervalu $(-1, 1)$, kde kladné hodnoty podporují diagnózu a záporné svědčí proti diagnóze. Schématicky je postup výpočtu míry podpory znázorněn na obrázku 6. K určení míry je použit:

1. Slovní popis znaku (např. teplota je vždy vysoká) pro diagnózu D
2. Důležitost znaku Z
3. Uživatelem zadaná slovní nebo číselná hodnota znaku, včetně jeho jistoty
4. Významové funkce

Expertní systém nejdříve určí míru shody popisu znaku (1) s popisem získaným od uživatele (3). U znaků, pro které jsou k dispozici významové funkce jednotlivých atomických termů, lze aplikovat jazykové operátory na významové funkce a následně určit jejich překryv. Míra shody se pohybuje v intervalu $\langle 0, 1 \rangle$.

Důležitost znaku může být zadána explicitně, nebo ji lze odvodit na základě příslovce častosti.

Kombinací důležitosti znaku, jistoty uživatele a míry shody popisů získáme míru v jaké znak podporuje diagnózu.

6.3.1 Seřazení diagnóz

Posledním úkolem je seřadit diagnózy na základě příspěvků jednotlivých znaků. Na začátku jsou diagnózy, které jsou nejvíce podobné popisu, na konci jsou ty, které se nejvíce liší.

Tento úkol není ještě zcela uspokojivě vyřešen. Možností, jak kombinovat příspěvky, je celá řada a posouzení výsledků jednotlivých metod může provést pouze expert. Báze znalostí je ve stádiu dokončování a posouzení vhodnosti jednotlivých metod již bohužel není možno z časových důvodů zahrnout.

6.3.2 Rozhraní pro volání vyhodnocovacích metod

Aby bylo možno testovat a kombinovat jednotlivé vyhodnocovací metody a současně nebylo nutné pracně měnit velkou část programu pokud chceme zaměnit jednu metodu za druhou, DGM definuje rozhraní pro volání těchto metod. Během vyhodnocování inferenční stroj prochází jednotlivé diagnostické jednotky a volá predikáty pro určení míry podpory diagnózy, kombinaci příspěvků jednotlivých znaků, srovnání dosaženého skóre a zobrazení výsledků. Programátor, který chce například implementovat novou metodu pro určení příspěvku znaku, nemusí znát ani strukturu báze znalostí ani to kde a jak jsou uloženy informace získané od uživatele. Stačí, aby znal pouze aritu a typy argumentů predikátu, který je volán během vyhodnocování. Potom je potřeba “pouze” napsat nový predikát a oznámit DGM jméno predikátu, který má volat (export jména).

Konverze jistoty uživatele

export: fcset_abc
arita: 2
argumenty: (+JU,-UBCres)

Převádí slovně zadanou uživatelskou jistotu JU na interval $\langle 0, 1 \rangle$

Příklad:

```
my_abc(sure, 0.7).
my_abc(maybe, 0.3).
:-fcset_abc(my_abc).
```

Předzpracování popisu znaku

export: fcset_sdc
arita: 5
argumenty: (+TypZnaku,+PopisZnaku,+Důležitost,FuzzyP,-SDCres)

Tento predikát konvertuje slovní popis znaku a jeho důležitost do vhodnějšího tvaru. DGM může využít tento mezistupeň při předzpracování báze znalostí (například se zde může provést syntaktická analýza textu a úprava významových funkcí). *FuzzyP* je fuzzy popis atomických termů. *SDCres* je výsledek konverze.

Určení míry podpory diagnózy D znakem Z

export: fcset_comb
arita: 5
argumenty: (+SDCres,+PopisZnakuUživatелеm,+UBCres,-SPres)

Na základě předzpracovaných znalostí *SDCres* a popisu znaku uživatelem je stanovena míra podpory diagnózy *SPres*

Kombinace příspěvků jednotlivých znaků

export: fcset_dgcc
arita: 3
argumenty: (+D,+SPresList,+REFres,-DGCCres)

SPresL je seznam [(IdZnaku,SPres)]..., který obsahuje příspěvky jednotlivých znaků pro diagnózu *D*. Argument *REFres* zatím není využit. Výsledkem je term (skóre) *DGCCres* který, popisuje jak se shoduje popis diagnózy ve znalostní bázi s popisem uživatele.

Porovnání

export: fcset_cmp
arita: 3

argumenty: (+DGCCres1,+DGCCres2,-CMPres)

Tento predikát je volán při seřazování diagnóz podle získaného skóre. Porovná termy *DGCCres1* a *DGCCres2*. *CMPres* může nabývat hodnot $<$, $>$, $=$.

Zobrazení výsledků

export: fcset_evrtt

arita: 2

argumenty: (+DGCCres,-Atom)

Tvar *DGCCres* termu není specifikován a proto je nutné definovat jeho konverzi na textový řetězec, který bude zobrazen.

Příklad:

```
num_cmp(N,N,'=').
num_cmp(N1,N2,'<'):- N1 < N2.
num_cmp(N1,N2,'>'):- N1 > N2.

to_text(Term,Txt):-float(Term),float_text(Term,Txt,fixed(8)).
to_text(Term,Txt):-atomic(Term),Txt=Term.

my_abc(sure,0.7).
my_abc(maybe,0.3).
my_sdc(A,B,C,D, A:B:C:D).
my_comb(_,_,_,_,RS):-rnd(RS).
my_dgcc( RL, _, Res ):-
    sum(RL,Res)
sum([],0).
sum([(_,CMB)|T],Sum):-
    sum(T,S2),
    Sum is CMB + S2.

:-fcset_all(my_abc,my_sdc,my_comb,my_dgcc,num_cmp,to_text).
```

Výše uvedený příklad je samozřejmě velmi zjednodušen. Míra podpory diagnózy je generována náhodně. Skóre je součet všech příspěvků.

Predikát *fcset_all/6* exportuje jména všech šesti predikátů. V případě volné proměnné na místě libovolného argumentu je ponecháno původní nastavení daného jména.

6.3.3 Použití významových funkcí

DGM poskytuje následující predikáty, pro práci s významovými funkcemi.

apply_lo(+JazykovýOperátor,+ f_1 ,- f_2) aplikuje jazykový operátor na významovou funkci f_1 . Výsledkem je modifikovaná významová funkce f_2 .

Důležitým úkolem při vyhodnocování je určit nakolik popis znaku, tak jak ho zadal uživatel, odpovídá popisu znaku v diagnostické jednotce. Nechť $d \in D$ je term, který popisuje znak podle subjektivního dojmu uživatele. $p \in P$ je popis znaku (term) v diagnostické jednotce. f_d a f_p jsou jejich významové funkce, aproximované pomocí úseček. Definujeme

$$S : D \times P \rightarrow \langle 0, 1 \rangle$$

jako

$$S\langle d, p \rangle = \frac{M(f_d \cap f_p)}{M(f_d)}$$

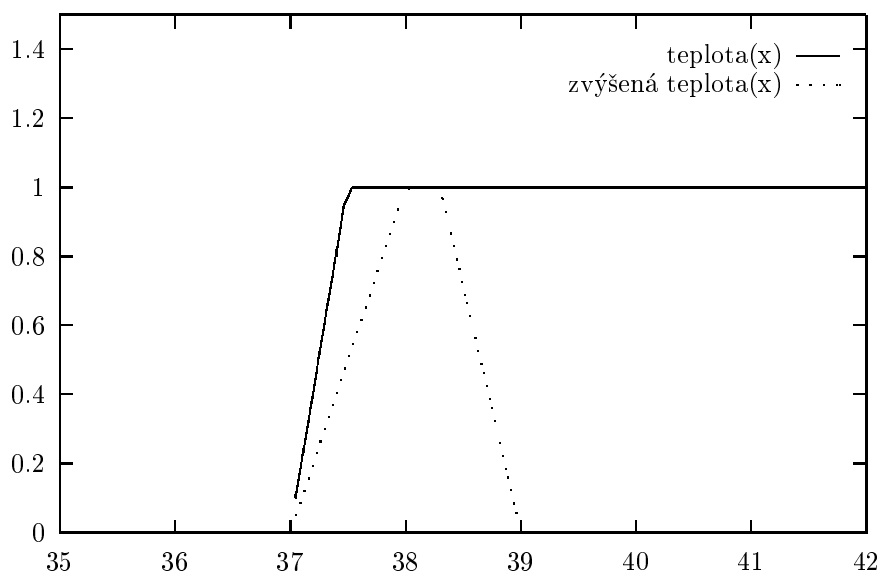
kde $M(f)$ je plocha ohraničená osou x a funkcí f . $S\langle d, p \rangle$ je stupeň, udávající jak term d splňuje popis p . K výpočtu tohoto stupně lze použít predikát **shs_coef**(+ f_d ,+ f_p ,-Coef). Uvedeme následující příklad: nechť $d =$ 'zvýšená teplota' a $p =$ 'pacient má teplotu'. Významové funkce $f_{teplota}$ a $f_{zvýšená\ teplota}$ jsou znázorněny na obrázku 7. Potom $S\langle d, p \rangle = 1$. Pokud by situace byla opačná, tj. $d =$ 'pacient má teplotu' a $p =$ 'pacient má zvýšenou teplotu', potom $S\langle d, p \rangle < 1$.

7 Závěr

Expertní systém napodobuje uvažování klasifikujícího diagnostika tam, kde je možno standartním způsobem popsat jednotlivé diagnostické jednotky. Je vhodný pro patologii, rtg a jiné klasifikační obory.

V patologii existují dva základní diagnostické problémy: klasifikační a gradační. Při klasifikaci jde o určení diagnózy, u gradačního problému jde o zařazení diagnózy na ose mírná — střední — těžká. Uvedený expertní systém je zaměřen spíše na problém klasifikační.

Při tvorbě expertního systému byl kladen důraz na jednoduchost plnění báze znalostí za použití medicínské literatury, která systematicky popisuje jednotlivá onemocnění. Expertní systém dále umožňuje popis vágních pojmů pomocí fuzzy množin a použití jazykových operátorů.



Obrázek 7: Významové funkce $f_{teplota}$ a $f_{zvýšená\ teplota}$

Diagnostika je dále podpořena modulem pro diferenciální diagnostiku a kontextovou nápovědou k jednotlivým znakům a topografiím s komentovanými obrázky jednotlivých onemocnění.

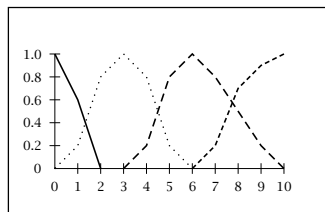
V současné době je naplňována báze znalostí z oblasti kožních onemocnění a doladován inferenční mechanismus.

Literatura

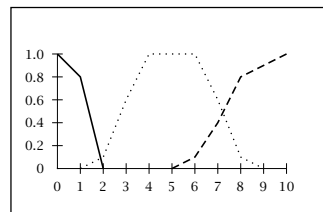
- [1] Novák V.: Fuzzy množiny a jejich aplikace, Matematický seminář SNTL, vol. 23, SNTL Praha, 1986.
- [2] Liebowitz J., De Salvo D. A., editors: Structuring Expert Systems, Prentice-Hall, 1989
- [3] Feit J., Matyska L.: Prolog Expert System for Liver Biopsies
- [4] Zadeh L. A.: Fuzzy Sets, Information and Control 8, 1965
- [5] Zadeh L. A.: Quantitative Fuzzy Semantics
- [6] Zadeh L. A.: The concept of Linguistic Variable and its Application to Approximate Reasoning, Information Sciences 8, 1975.
- [7] Zadeh L. A.: Test–Score Semantics as a Basis for a Computational Approach to the Representation of Meaning, University of California, Berkeley, USA.
- [8] Kulka J.: Aproximativní kvantifikace v psychologii — Lingvistický přístup, Sborník prací Filozofické Fakulty MU, 1980.
- [9] Adlassnig K. P.: Fuzzy Set Theory in Medical Diagnosis, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-16, No. 2, 1986.

A Definice významových funkcí pro vybrané znaky

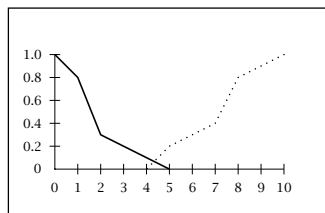
infiltrate_density:
none, slight, moderate, dense



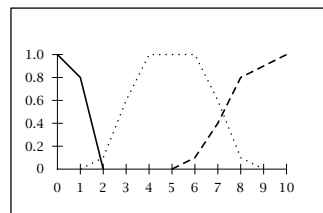
neutrophils:
none, admixture, dominate



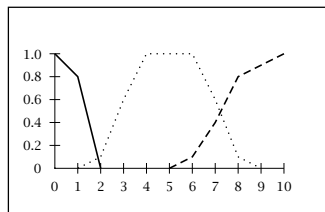
mast_cells:
none-or-few, many-or-dominate



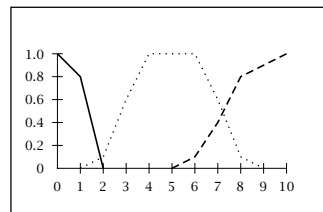
melanophages:
none, scattered, many



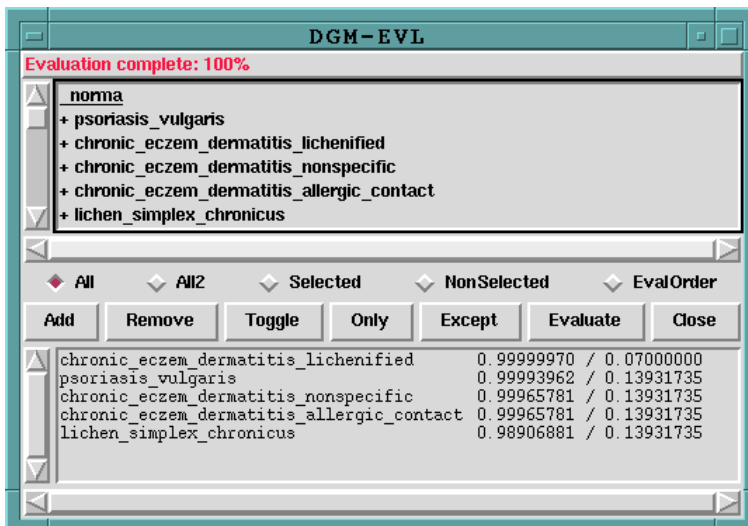
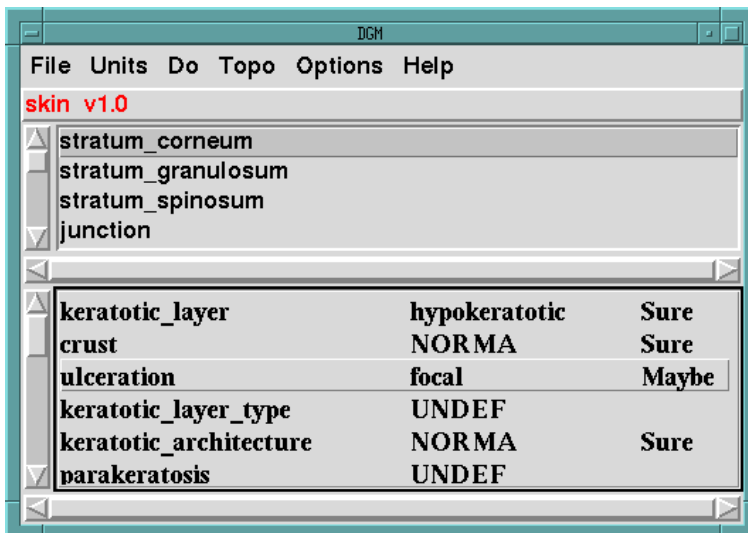
multinucleated_cells:
none, scattered, many



munro_microabscesses:
none, occasional, many



B Ukázka grafického rozhraní



C DGM Help

DGM umožňuje definovat nápovědu pro každou topografii, diagnostický znak, popř. globální nápovědu pro danou znalostní bázi. Momentálně jsou k dispozici následující tři predikáty pro tvorbu nápovědy:

- **:-help_unit(tcl_skript)**
- **:-help_topo(topografie, tcl_skript)**
- **:-help_sign(topografie, znak, tcl_skript)**

Nápověda je aktivována po stisknutí klávesy *<Help>*. V závislosti na aktuálním kontextu DGM okna a pozici kurzoru je interpretován příslušný tcl skript. Díky použití tcl je nápověda velmi flexibilní. Tcl skript může obsahovat příkazy ke spuštění dalších aplikací, prohlížečů obrázků, textu, videa atp.

Dá se však předpokládat, že nejvíce bude požadováno zobrazování textu a obrázků s komentáři. Proto jsem v Tk napsal hypertextový prohlížeč, který je součástí DGM.

C.1 DGM browser

X okno DGM browseru sestaveno ze čtyř částí (viz obr 8):

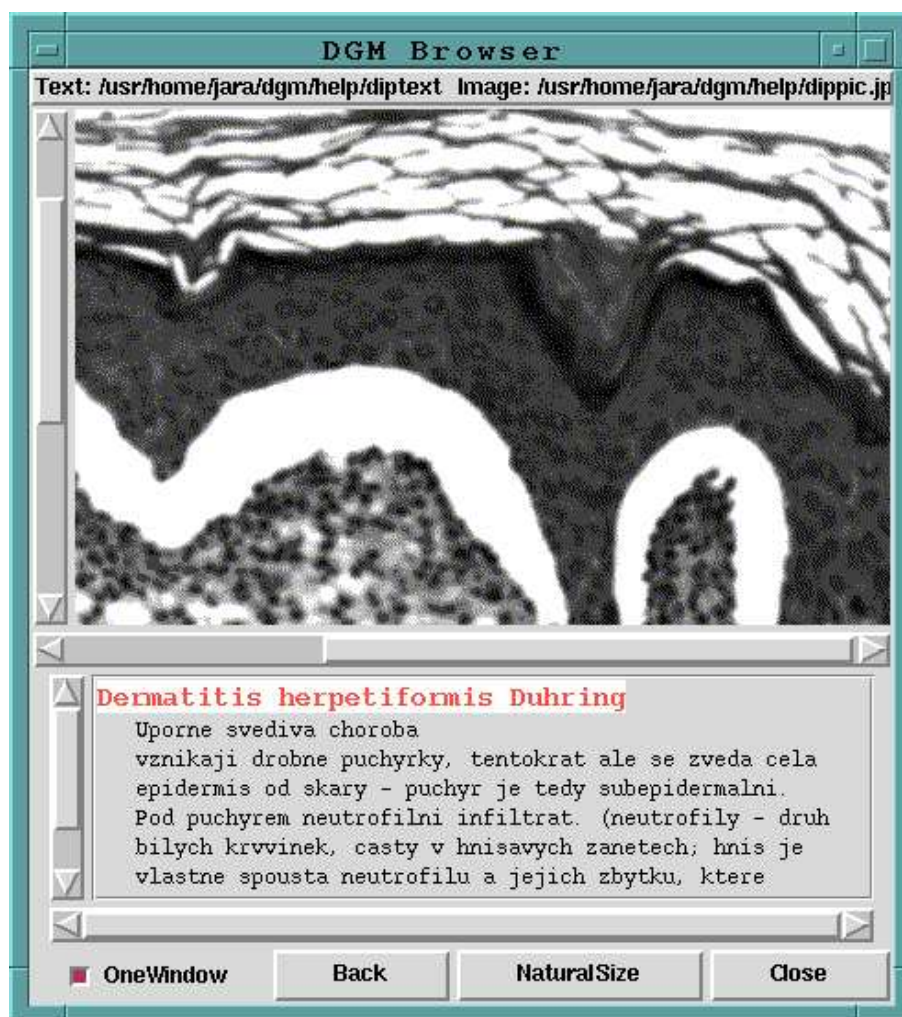
- Horní lišta je defaultně používána k zobrazení jména právě zobrazovaného souboru.
- V případě, že je požadováno zobrazení obrázku, pak je pro něj vytvořena oblast pod horní lištou a nad textovým přípravkem (widgetem). Další částí je oblast textu. Text může obsahovat hypertextové odkazy, a jeho vzhled může být ovlivněn mnoha způsoby.
- Ve spodní části se nachází příkazová lišta:

OneWindow Pokud není zatrženo, je vždy pro zobrazení odkazu vytvořeno nové okno. Jinak je odkaz zobrazen v aktuálním okně.

Back DGM browser zaznamenává historii průchodu po daných odkazech. Back vrací DGM o jeden odkaz zpět.

NaturalSize Změní velikost okna tak, aby obsahovalo celý obrázek + text požadované výšky a šířky.

DGM browser poskytuje následující tcl/tk procedury pro zobrazení textu či grafiky:



Obrázek 8: DGM browser

- **n_{text} r_{text} d_{text}** pro zobrazení textu. První znak v názvu udává, kam bude text zobrazen. **n_{text}** vytvoří pro text nové okno, **r_{text}** v případě, že je volána v rámci kliknutí myši na hypertextový odkaz, přepíše původní text, který tento odkaz obsahoval. Pokud je volána v reakci na klávesu *<Help>*, pak je text zapsán do “DGM default help” okna. **d_{text}** se chová jako **r_{text}** nebo **n_{text}** a to v závislosti na hodnotě OneWindow “checkboxtonu”. První povinný argument udává jméno souboru, ze kterého bude text načten. Pokud je prvním znakem jména #, zobrazí se přímo text následující po znaku #. Druhý nepovinný argument udává, co bude zobrazeno v horní liště.
- **npic rpic dpic** pro zobrazení grafiky. Platí pro ně vše, co bylo řečeno o zobrazení textu. DGM browser umí zobrazit formáty PPM a GIF. Pro zobrazení JPG musí být v systému dostupný **djpeg** a název souboru musí mít příponu .jpg
- **npictext rpictext dpictext** pro zobrazení grafiky a textu. První argument je jméno souboru obsahujícího obrázek, druhý udává text, který má být zobrazen (viz. procedury pro zobrazení textu), poslední nepovinný obsahuje text pro horní lištu.

C.2 Tvorba textu pro DGM browser

Text zobrazovaný pomocí procedur **n_{text} r_{text} d_{text} npictext rpictext dpictext** může obsahovat příkazy pro formátování a vytvoření hypertextových odkazů. Každý příkaz začíná obráceným lomítkem, následuje jeho název a případné argumenty uzavřené ve složených závorkách. Nejsou dovoleny mezery mezi jménem a první složenou závorkou (*\textopt_{}*) a mezi *{} (\link{}_{}*).

\tag{text}{tag_list} na text uvedený v prvním argumentu jsou aplikovány vlastnosti uvedené v argumentu druhém. V tk jsou tyto “vlastnosti” označovány jako **tags**. **tags** jsou symbolické názvy definované v tk, nebo příkazem *\newtag.tag_list* je seznam “vlastností” oddělených mezerami.

Příklad: *\tag{foo text}{big underline}* zobrazí “foo text” zeleně, s podtržením a fontem adobe velikosti 110pt - viz. definice vlastnosti **big** níže.

\link{text}{tcl_skript}{tag_list} vytvoří hypertextový odkaz. Pokud se kurzor myši bude nacházet nad tímto textem, bude změněn na “hand2”,

a změni se barva textu. Po kliknutí na text bude proveden *tcl_skript*. Třetí argument je nepovinný. Pokud není uveden, text je zobrazen s podtržením. Může být i prázdný.

Příklad: `\link{picture of}{npic $HP/pic.gif}{big}` vytvoří odkaz, který po nakliknutí myši zobrazí obrázek `pic.gif`.

`\defs, \enddefs` text uzavřený mezi těmito příkazy není zobrazován, skupinu níže uvedených příkazů je vhodné uzavřít mezi `\defs` a `\enddefs`.

`\newtag{tag_nameoptions}` definuje novou vlastnost textu.

Příklad: `\newtag{big -font -adobe-*-r-normal-*-110-*-*-*-*
-foreground green -relief raised}` definuje vlastnost textu nazvanou jménem **big**.

`\textopt{options}` konfiguruje textové okno. Volby (“options”) jsou detailně posány v dokumentaci tk (soubor `doc/text.n`).

Příklad: `\textopt{-height 10 -cursor hand1 -tabs{1c 3c 5c}}` nastaví výšku textového okna na 10 řádků, změni tvar kurzoru myši na **hand1** a nastaví tři tabulační pozice.

`\eval{tcl_skript}` provede daný `tcl_skript`.

DGM browser definuje globální proměnou HP, kterou je možno se odkazovat na kořenový adresář obsahující help soubory.