

E-MODELOVÁNÍ ZNALOSTÍ PRO PODPORU ROZHODOVÁNÍ

E-KNOWLEDGE MODELLING IN A DECISION PROCESS

David Buchtela

Anotace:

Dnešní hranice konkurence, se více než kdy jindy, usadila v kreativitě a schopnostech, expertíze a dovednostech, zlepšování a inovacích. Proto je v dnešním tržním prostředí, kdy jedinou jistotou je nejistota, jediným jistým zdrojem přetrvávající konkurenceschopnosti znalost. Úspěšné společnosti proto trvale vytváří nové znalosti, které rozšiřují v rámci celé organizace, ztělesňují je do nových výrobků a průběžně inovují.

Pro naše podnikové prostředí je čím dále tím více typické zahlcení podniku informacemi, přičemž těch pravých a požadovaných informací je přitom k dispozici příliš málo, často zcela chybí.

Je rozdílné pracovat s informacemi a znalostmi. Informace snižuje míru neurčitosti u příjemce, říká mu něco, co neznal. Znalost vzniká zasazením množství informací do kontextu a nalezením jejich kauzálních vazeb. Znalosti jsou založeny na kontextové hodnotě informací. A tady se narazilo na problémy v současných informačních systémech, jejichž databáze mají velmi omezené možnosti pro zachycení složitějšího kontextu. Významná část informačního a znalostního potenciálu by však měla být u jednotlivců, organizací, států či jiných skupin uchováвана v digitální podobě a tvořit znalostní bázi podniku, oboru apod..

Získané znalosti v různých oborech je možné formalizovat do doporučení (*guidelines*) usnadňující rozhodování v daném konkrétním případě, např. rozhodování lékaře léčícího pacienta na základě doporučení opírajících se o znalostní bázi oboru. Pro počítačovou implementaci a zpracování je nutné mít tato doporučení ve strukturované formě.

Pro přehlednou reprezentaci doporučení slouží model GLIF (*GuideLine Interchange Format*), který vznikl spoluprací Univerzity Columbia, Harvardské, McGillovy a Stanfordské univerzity. GLIF poskytuje objektivě a procesně orientovaný pohled na doporučení. Výsledným modelem je orientovaný graf skládající se z pěti hlavních částí: *akce* představující specifickou činnost nebo událost; *rozhodování* představující větvení na základě automatického splnění logického kritéria nebo rozhodnutí uživatele; *rozdělení* a *synchronizace* – rozdělení se používá při modelování nezávislých kroků, které mohou probíhat paralelně a synchronizace slouží pro tyto kroky jako slučovací bod; *stav* značící stav, ve kterém se zkoumaný objekt nachází.

Získaný formalizovaný model je možné zabudovat, implementací ve vhodném programovacím jazyku (např. UML, XML, ...), do znalostních informačních systémů a použít pro podporu i kontrolu rozhodování v souladu s expertními doporučeními příslušného oboru.

Abstract:

A present limit of competition is consisted by creativity, abilities, expertise, crafts and improvements more than other times. That is why knowledge is the only source of surviving competitive ability in marketing environment where uncertainty is the only certainty. Therefore successful companies permanently create new knowledge which they extend in whole company, put in new products and improve them all the time.

Informative glut is typical for our companies more and more. There are so little correct and requisite information there, they are often missing.

It is different to work from information and knowledge. The information reduces recipient indefinite rate. It say something unknown to him. Knowledge is caused by putting quantity of information in a context and finding their causal structure. Knowledge is found on context worth of information. There are problems in current information systems here which databases have very pure possibility of putting complicated context down. Significant part of information and knowledge potential of individual, companies, states or other groups should be kept in digital form and create knowledge base of company, branch etc.

Knowledge acquired in different branches is possible to represent by guidelines which make decision process in concrete cause easy, for example doctor decisions in patient treatment which are in harmony with clinical guidelines. For computer implementation and processing it is necessary to have guidelines explicitly structured.

The most important and nowadays mostly used is the GLIF (Guideline Interchange Format) model. The GLIF model is result of collaboration among Columbia University, Harvard University, McGill University and Stanford University. GLIF favours the objected and process-oriented view of guidelines. GLIF is mostly given as a flowchart representing a sequence of steps. Action steps specify work to be performed. Decision steps represent decisions by automatic evaluating logical criteria or choices that should be made by the user. Branch steps are used to model concurrent guideline steps. All of these guideline steps can occur in parallel. Synchronisation steps are used in conjunction with branch steps. State steps characterize a object state.

Final GLIF model can be mount into knowledge information systems by implementation in useful programming language (for example UML, XML,...) and be used to support and controlling of decision process in harmony with branch expert guidelines.

Klíčová slova:

Znalosti, znalostní systém, GLIF model

Key words:

Knowledge, knowledge systems, GLIF model

Úvod

Pro naše podnikové prostředí je čím dále tím více typické zahlcení podniku informacemi, přičemž těch pravých a požadovaných informací je přitom k dispozici příliš málo, často zcela chybí. Přitom konkurenceschopné podniky jsou dnes převážně ty, které se vyznačují zejména prozíravostí a předvídavostí vývoje, mobilizace a alokace svých znalostních zdrojů. Právě rozvoj v této oblasti a péče o tyto zdroje se stávají v současnosti nejdůležitějšími determinanty jejich udržitelnosti a vysoké výkonnosti.

Cíl

Tento článek vznikl za účelem seznámení se s problematikou získávání a reprezentace znalostí v expertních systémech. Podrobněji se zabývá formalizováním znalostí (oborových doporučení), jejich grafickou reprezentací modelem GLIF a implementací v jazyce UML a XML.

Metodika

Informace versus znalosti

Znalostmi obecně rozumíme jakékoliv informace o reálném světě. Získáváme je poznáváním jednak přímo - pozorováním objektů a jevů reálného světa, jednak nepřímo - odvozováním. Primární je vždy přímé pozorování. Každé získávání znalostí je vázáno na jazyk, kterým vypovídáme o té části světa, kterou poznáme. V běžném životě obvykle používáme k vyjádření svých znalostí přirozeného jazyka. Pro popis úzce vymezeného oboru poznávání skutečností se používají umělé jazyky (matematické formule, chemické značky, programovací jazyky, jazyk schémat elektrických obvodů atd.), které jsou jednoznačnější a úspornější.

Je rozdílné pracovat s informacemi a znalostmi. Informace odstraňuje míru neurčitosti u příjemce, říká mu něco, co neznal. Je lehce přenositelná pomocí technologie na jiného příjemce. Znalost vzniká zasazením množství informací do kontextu a nalezením jejich kauzálních vazeb. Znalosti jsou založeny na kontextové hodnotě informací.

Komplexní poznání (někdy dokonce označované pojmem moudrost) je pak možno považovat za množinu znalostí, informací a dat vztahujících se k určité problematice.

Znalostní systém

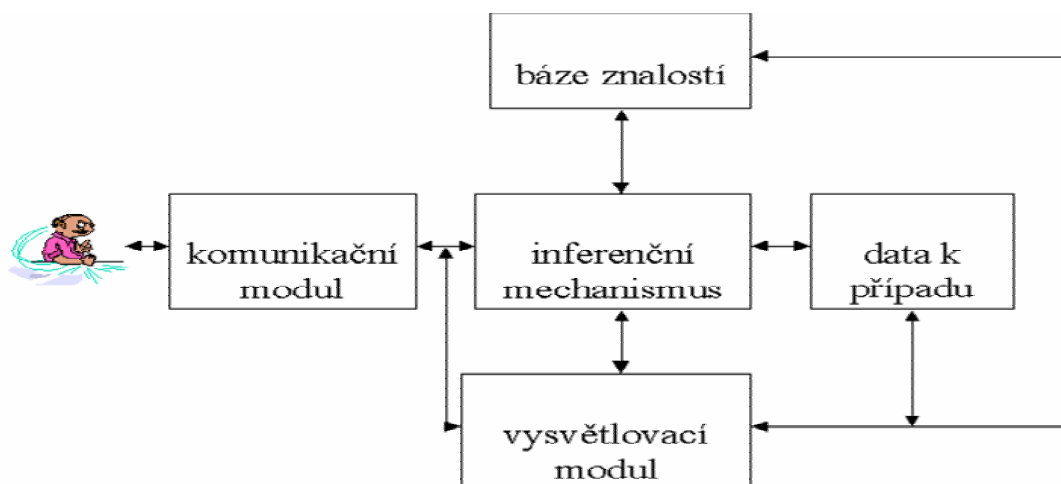
Znalosti lze posuzovat podle tvaru, ve kterém jsou reprezentovány. Můžeme je rozdělit na:

- znalosti reprezentované **deklarativně**, které vyjadřují, co je nebo má být poznáno resp. dokázáno
- znalosti reprezentované **procedurálně**, které říkají, jak poznávat nebo odvozovat

Deklarativně reprezentované znalosti budeme nazývat **poznatky**. Procedurálně reprezentované znalosti budeme nazývat **pravidla**.

Tatáž znalost může být často reprezentována buď jako poznatek ("Pes je zvíře") nebo jako pravidlo ("Je-li X pes, potom je X zvíře"). Soubor všech jedinečných znalostí o daném případě nazýváme **báze dat** (dále data). Soubor všech obecných znalostí týkajících se dané problematiky nazýváme **báze znalostí** (*knowledge base*). Většinou jsou data reprezentována deklarativně a znalosti procedurálně. Znalosti získáme od experta a při řešení úlohy jsou aplikovány na daná data

Znalostní systém je realizovaný poměrně rozsáhlou soustavou kooperujících programů. Jednotlivé programové celky této soustavy tvoří prvky funkčně vymezených a svou strukturou i posláním odlišných modulů. Vezmeme-li do úvahy i jejich vzájemné vazby, tvoří tyto moduly **architekturu znalostního systému**.



Obr. 1: Architektura znalostního systému

V každém tradičně koncipovaném znalostním systému je možné rozlišit čtyři základní složky, a to **bázi znalostí**, **bázi dat**, **inferenční mechanismus** a **komunikační modul**.

Báze znalostí

Báze znalostí reprezentuje všeobecně platné a přijímané poznatky o pravidlech a zákonitostech z vymezené oblasti. V průběhu řešení se k nim přistupuje zřídka, někdy pouze jednorázově. Významnou charakteristikou báze znalostí je i to, že se v průběhu procesu nemění.

Báze dat

Báze dat je nositelkou konkrétně zadaných nebo odvozených faktů, popřípadě předpokládaných (až odhadnutých) údajů o nějakém specifickém problému. Jsou to informace, které vychází z konkrétního prostředí, konkrétních zařízení a provozů. Báze dat tvoří tu část expertního systému, která nejvíce odráží údaje experta a jeho praktické zkušenosti. Do báze dat se také na počátku vkládají vstupní údaje, během procesu se zde ukládají odvozená fakta, mezivýsledky a výsledky.

Inferenční mechanismus

Inferenční mechanismus je složen ze souboru kooperujících programů zabezpečující procedurální složku činnosti expertního systému. Tím inferenční modul umožňuje v určitém rozsahu napodobovat expertovu schopnost uvažovat. Modul simuluje především ty schopnosti, které souvisí s efektivním využíváním poznatků a zkušeností, získaných na základě asociací, hierarchií, příčinně-důsledkových vazeb, kontextů a spojování poznatků do vhodně souvisejících celků a posloupností.

Komunikační modul

Znalostní systém většinou není izolovaný od ostatního (softwarového nebo fyzického) prostředí a musí si tedy s ním vyměňovat informace, minimálně musí uživatel zadat otázku nebo problém, který se má znalostním systémem zpracovávat. Úlohu zadávání a zpracovávání údajů z vnějšího okolí zabezpečuje modul styku s prostředím.

Získávání znalostí z databáze

S aparátem, jako jsou databázové systémy (relační i jiné), se v oblasti umělé inteligence zpravidla nevystačí, takže se pozornost soustřeďuje právě na soubory zápisu znalostí, a to jak z hlediska teoretického, tak z hlediska aplikací. Různé aplikace samočinných počítačů ve výzkumu, administrativě, ve výrobě atd. jsou založeny na tom, že je třeba ukládat do paměti počítačů velmi rozsáhlé soubory údajů se složitými vzájemnými vztahy, podle potřeby jednotlivé údaje z paměti vyjímat, doplňovat a upravovat. Zároveň systém reprezentace znalostí musí umožnit, aby na základě uložených údajů mohly být vyvozeny příslušné důsledky, asi tak, jako člověk vyvozuje důsledky ze sdělených informací.

Data mining

Proces objevování znalostí, pomocí získávání dříve neznámých akčních informací z velmi rozsáhlých databází se nazývá data mining.

Proces data miningu lze srovnat s hledáním úlomků zlata, avšak velmi sofistikovaným geologickým průzkumem. Při provádění data miningu nalézáme skryté vzory a trendy v rozsáhlých databázích, které pak slouží jako podklady pro rozhodování.

Některé definice data miningu jsou založeny na specifických analytických technikách jako např. neuronové sítě, genetické algoritmy apod. Pro různé úlohy a data se hodí různé metody, např. indukce, dedukce, redukce, analogie, generování a testování, abdukce nebo heuristická pravidla. Velmi častý je případ, kdy nejlepších výsledků dosáhneme vhodnou kombinací různých metod.

Reprezentace znalostí

Na reprezentaci znalostí jsou kladeny často protichůdné požadavky. Při vytváření báze znalostí je velmi důležité, aby použitá reprezentace dovozovala jednoduše upřesňovat znalosti. Zejména je důležité, aby bylo možné inkrementálně rozšiřovat bázi znalostí. Jde tedy o požadavek modularity reprezentace. Je zřejmé, že báze znalostí bude modulární, bude-li složena z nezávislých, jednoduchých, uniformních částí. Přidáním dalších takovýchto částí zjemníme bázi znalostí, vzhledem k vzájemné nezávislosti nezasáhne úprava jedné části zbytek báze atd. Na druhé straně, aby bylo možné jednotlivé znalosti rychle z báze vyvolávat, je důležité, aby příbuzné znalosti byly volně sdružovány a nebyly reprezentovány nezávisle.

Příkladem modulární reprezentace znalostí jsou produkční pravidla a jazyk predikátové logiky 1. řádu, myšlenku sdružování podobných znalostí vystihují rámce a scénáře a sémantické sítě.

Formalizace znalostí

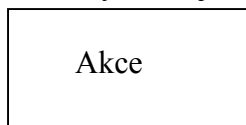
Získané znalosti v různých oborech je možné také formalizovat do doporučení (*guidelines*) usnadňující rozhodování v daném konkrétním případě, např. rozhodování lékaře léčícího pacienta na základě doporučení opírajících se o znalostní bázi oboru. Pro počítačovou implementaci a zpracování je nutné mít tato doporučení ve strukturované formě.

Výsledky a diskuze

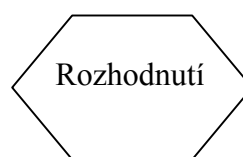
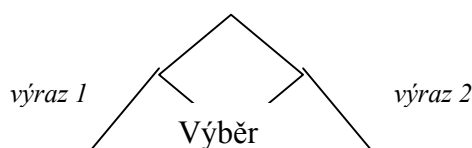
Model GLIF

Pro přehlednou reprezentaci doporučení slouží model GLIF (*GuideLine Interchange Format*), který vznikl spoluprací Univerzity Columbia, Harvardské, McGillovy a Stanfordské univerzity. GLIF poskytuje objektivě a procesně orientovaný pohled na doporučení. Výsledným modelem je orientovaný graf skládající se z pěti hlavních částí:

- **Akce** (*Action step*) - představuje specifickou činnost nebo událost. Akcí může být i podgraf, který dále zjemňuje danou činnost.



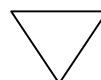
- **Rozhodování** (*Decision step*) – představuje větvení (výběr) na základě automatického splnění logického kritéria (*Case step*), kdy další postup grafem je dán výsledkem aritmetického nebo logického výrazu nad konkrétními daty, a nebo rozhodnutí uživatele (*Choice step*), kdy akce reprezentovaná tímto krokem může ale nemusí být provedena, nebo může probíhat paralelně s jinými akcemi. Uživatel má v tomto místě možnost se rozhodnout, kterou částí grafu bude dále pokračovat.



- **Větvení a synchronizace** – větvení (*Branch*) se používá při modelování nezávislých kroků, které mohou probíhat paralelně a synchronizace (*synchronisation*) slouží pro tyto kroky jako slučovací bod.

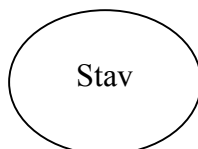


Větvení



Synchronizace

- **Stav** (*state step*) – značí stav, ve kterém se zkoumaný objekt nachází po provedení předchozího kroku.



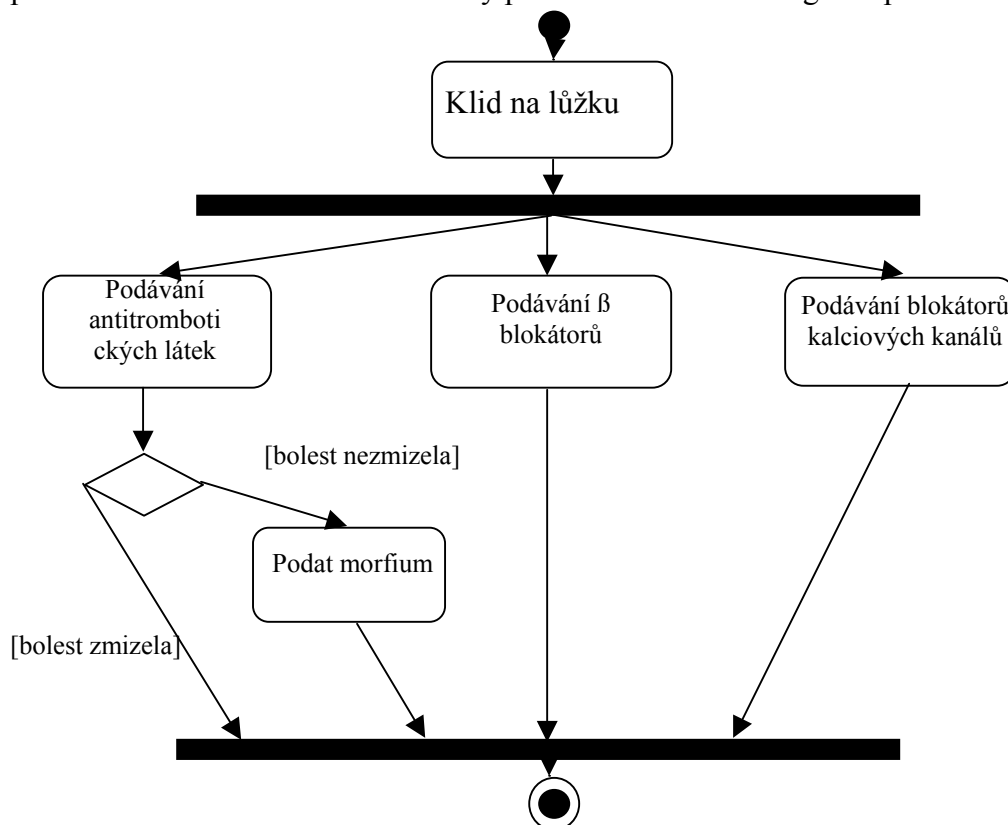
Získaný formalizovaný model je možné zabudovat, implementací ve vhodném programovacím jazyku (např. UML, XML, ...), do znalostních informačních systémů a použít pro podporu i kontrolu rozhodování v souladu s expertními doporučeními příslušného oboru.

Zobrazení modelu GLIF pomocí UML

Pro grafické zobrazení modelu GLIF pomocí UML (*Unified Modeling Language*) se jako vhodný prostředek nabízí diagram aktivit. Následující tabulka uvádí možné převedení jednotlivých částí modelu GLIF na stavební kameny diagramu aktivit.

GLIF	UML
Akce (Action step)	Aktivita
Rozhodování (Case step)	Rozhodovací blok
Výběr (Choice step)	Aktivita následovaná rozhodovacím blokem
Stav (State step)	Stav Počáteční a koncový bod
Větvení (Branch)	Synchronizační bod
Synchronizace (synchronisation)	Synchronizační bod

Jako příklad uveďme model základní léčby pacienta s nestabilní anginou pectoris



Obr. 3: model základní léčby v UML

Implementace GLIF modelu pomocí XML

Získaný GLIF model je možné, pro účely dalšího počítačového zpracování, uložit ve formátu XML (*eXtensible Markup Language*), který je v současné době přehledným a rozšířeným způsobem uložení veškerých dat v počítačových (nejen expertních) systémech. Následující výpis ukazuje část možné reprezentace výše uvedeného příkladu v XML.

```

<?xml version="1.0" encoding="UTF-8"?>
<Step>
  <name>Step1</name>
  <type>action</type>
  <caption>Akce</caption>
  <text>&lt;H&gt;Klid na lůžku&lt;/H&gt;</text>
  <tag/>
  <x>300</x>
  <y>1</y>
  <w>200</w>
  <h>70</h>
  <focus>1</focus>
  <status>0</status>
  <next>
    <F1>&lt;Step2&gt;:(400,71),(400,100)</F1>
  </next>
</Step>
<Step>
  <name>Step5</name>
  <type>choice</type>

```

```

    <caption>Rozhodnutí</caption>
    <text>&lt;H&gt;Podávání&lt;/H&gt;&lt;H&gt;antitrombotických&lt;/H&gt;&lt;H&gt;láték&lt;/H&gt;</text>
    <tag>NAPrevize2002.htm#antitrombotika</tag>
    <x>500</x>
    <y>150</y>
    <w>200</w>
    <h>100</h>
    <focus>1</focus>
    <status>0</status>
    <next>
      <F1>&lt;Step11&gt;:(600,250),(600,530),(10,530),(10,700),(400,700),(400,720)</F1>
    </next>
  </Step>
  <Step>
    <name>Step6</name>
    <type>case</type>
    <caption>Výběr</caption>
    <text>&lt;H&gt;Bolest zmizela?&lt;/H&gt;</text>
    <tag/>
    <x>10</x>
    <y>280</y>
    <w>280</w>
    <h>100</h>
    <focus>0</focus>
    <status>0</status>
    <next>
      <F1>&lt;Step11&gt;Ano:(60,380),(60,530)</F1>
      <F2>&lt;Step7&gt;Ne:(200,380),(200,420)</F2>
    </next>
  </Step>

```

Závěr

Reprezentace a modelování znalostí je dnes aktuálním tématem ve výzkumu věnovaném umělé inteligenci. Tam, kde nejde o řešení úkolů čistě formálních (dokazování teorémů apod.), ale o problémy mající empirický podklad (od her, jako jsou šachy, přes řízení výrobních a jiných složitých procesů až po soustavy automatického zodpovídání dotazů v přirozeném jazyce, ovládání robota apod.), je nutné vypracovat speciální soustavu reprezentace znalostí, na kterých mohou operovat systémy při řešení problémů, vyhledávání informací, rozboru pokynů s poměrně elementárními postupy.

Využitím formalizace znalostí, konkrétně v medicíně, a jejich reprezentací GLIF modelem se mimo amerických univerzit zabývají i informatici společně s lékaři v EuroMISE centru při AV ČR, kde je dosahováno příznivých praktických výsledků.