



Teoretické zdroje pro systémovou integraci

Systémová integrace je proces, jehož cílem je sjednocení vzájemně nekompatibilních částí celku, odstranění neregularit na vazbách mezi částmi celku a vytvoření homogenního prostředí. Abychom lépe jako systémoví integrátoři dokázali uplatňovat náš zájem spočívající v integraci systémů, je dobré se seznámit s teoriemi, které jsou dnes součástí rozsáhlých teorií moderní informatiky a systémového inženýrství.

Tomáš Vitvar

Na teoretické zdroje systémové integrace nahlížíme z pohledu gnoseologického čtyřúhelníku, tzn. procesu poznávání a vytváření obrazu o originálu, jazykových prostředků pro konstrukci obrazu, procesu zpracování obrazu a kvality dat. Navazující poznatky na kvalitu dat jsou v závěru tohoto příspěvku věnovány způsobům reprezentací znalostí.

Proces poznávání a vytváření obrazu o originálu

Moderní informatika je rozvíjena ve třech dimenzích: jazykové prostředky pro konstrukci

obrazu o originálu, fáze technologického procesu zpracování obrazu, kvalita obrazu. Vytvořený, transformovaný a implantovaný obraz je pak nástrojem pro ovládnutí originálu.

Základní ilustrací gnoseologie, tj. procesu poznávání a vytváření obrazu o originálu, je koncept gnoseologického čtyřúhelníku (obr. 1) vyjadřující relace mezi prvky:

- pozorovatel P, resp. uživatel;
- objekt O, jako pozorované, zkoumané prostředí, originál, který je v zájmu pozorovatele, uživatele;
- jazyk J, jako prostředek pro komunikaci a vyjádření údajů, dat o objektu;
- informace I, jako podmnožina možných jazykových konstruktů s konkrétním použitím.

Jde o vlastní informační soustavu, resp. obraz objektu jako nástroj pro ovládnutí originálu.

Jazykové prostředky pro konstrukci obrazu

Jazyk je základním nástrojem pro komunikaci, pro zachycení vlastností o objektu a jeho vztahů s okolím. Z celkové množiny všech jazyků je v moderní informatice uplatňována především teorie jazyků formálních, která zahrnuje definici jazyka a metajazyka. Takové definice slouží především pro zachycení obrazu o originálu a pro vzájemnou komunikaci, přeložitelnost mezi jednotlivými obrazy různých originálů.

Základní koncept jazyka

Základní koncept jazyka je tvořen trojicí definičních složek:

- abeceda – množina terminálních, dále nedělitelných symbolů;
- syntaktická pravidla – množina pravidel umožňující konstrukci složitějších, přípustných výrazů (syntaktických výrazů, konstruktů) z terminálních symbolů;
- sémantika (význam) přiřazena k jednotlivým syntaktickým konstruktům.

Podle povahy jednotlivých složek

jazyka je možné jazyky členit na

- Přírozené, kdy jednotlivé složky mají povahu dynamickou. U přirozených jazyků dochází k přiřazování významu jazykovým konstruktům návykem, tradicí, pokusem nebo omylem na základě společenské zkušenosti.
- Odborné, kdy jednotlivé složky jazyka jsou předem vymezeny a definovány na základě zadání z okolí, podle pravidel, norem nebo konvencí.
- Formální, kdy sémantika jazyka je vůči celku jazyka irelevantní nebo je odvozená z prvních dvou složek.

Formální jazyky a metajazyky

Teorie formálních jazyků dnes představuje rozsáhlou teoretickou disciplínu a patří mezi základy teorie informatiky nacházející uplatnění v různých oblastech (programovací jazyky, jazyky pro popis struktury dat, společných rozhraní systémů, komunikačních protokolů a další). Základním prostředkem pro popis všech jazykových konstruktů umělého jazyka je metajazyk. Jedním z možných metajazyků pro popis syntaktických struktur formálních, umělých jazyků je gramatika.

Gramatika je definována jako čtveřice:

- N je konečná množina neterminálních symbolů, neboli množina metajazykových proměnných, která se používá pro označení celků jazyka.
- T je konečná množina terminálních symbolů, neboli abeceda, nad kterou je definován jazyk.
- P je konečná množina pravidel, syntaktických pravidel, určující předpisy pro konstrukci řetězců jazyka.
- S je počáteční, startovací symbol gramatiky, $S \in N$.

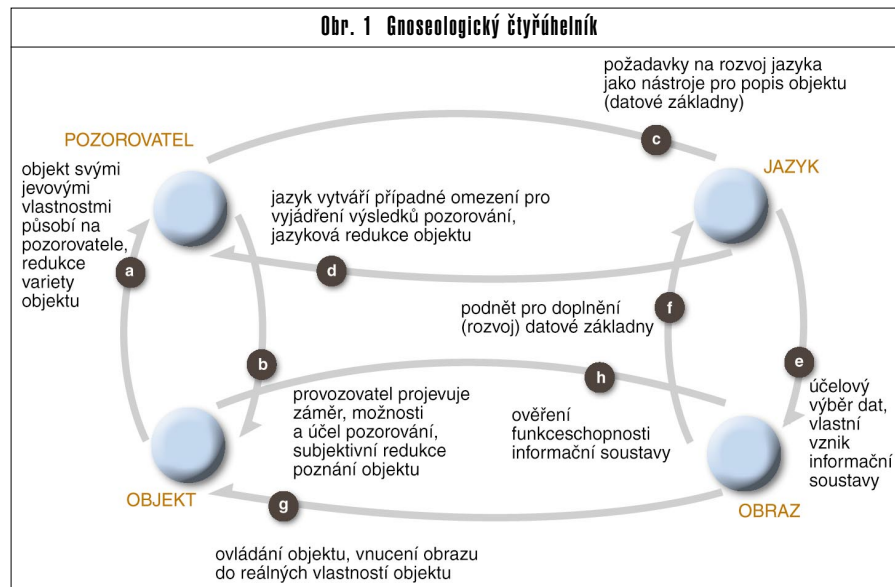
Gramatiky lze klasifikovat do čtyř tříd, a to podle tvaru syntaktických pravidel na gramatiky typu 0 (neomezené), typu 1 (kontextové), typu 2 (bezkontextové) a typu 3 (regulární). Pro jednotlivé třídy gramatik dále existují třídy automatů takové, že pro danou gramatiku lze sestavit automat, který popisuje stejný jazyk. Důsledkem vztahu automatů a gramatik je vytvoření algoritmu, syntaktického analyzátoru, který pro danou gramatiku umožňuje zjistit, zda libovolný řetězec je touto gramatikou generován, popř. strukturu tohoto řetězce.

Multijazyky a multijazykové prostředí

Východiskem pro analýzu multijazykového prostředí je gnoseologický čtyřúhelník (viz

obr. 1). Vznik multijazykového prostředí nastává při existenci více pozorovatelů P (uživatelů) resp. více jazyků J k jednomu pozorovanému objektu O . V rámci procesu zpracování obrazu tak dochází k vytváření vícenásobného obrazu o objektu a zároveň vzniká více různých jazyků jako prostředků

- Konceptuální schéma pojednává o uživatelských představách a potřebách vůči soustavě. Jejím cílem je zdokumentovat uživatelské požadavky a zachytit je ve standardní formě vhodné pro další zpracování.
- Logické schéma pojednává o transformaci konceptuálního schématu do jeho for-

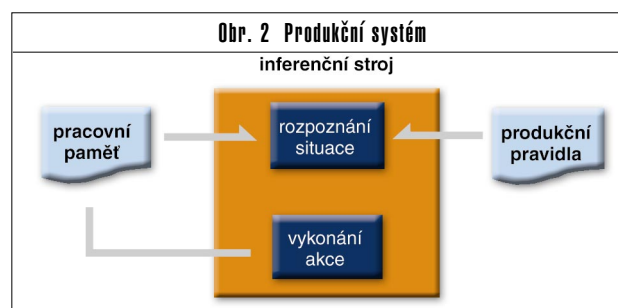


pro komunikaci a popis objektu. Takové jazyky jsou pak reprezentovány, generovány gramatikami se vzájemně odlišnými složkami množin abeced, pravidel a významů přiřazených k jazykovým konstruktům.

Při existenci více jazyků jako prostředků pro komunikaci, resp. více obrazů jako prostředků pro ovládání jednoho originálu, může docházet k formování celku s různorodými částmi, ke vzniku heterogenního prostředí. Různorodé části tvořící spojený celek jsou vybaveny předpoklady pro homogenizaci celku. Nalezení, vytvoření společného rozhraní mezi různorodými částmi systému vede ke zvyšování míry kompatibility a je základem pro vzájemné porozumění, spolupráci heterogenních částí jako předpokladu pro homogenní chování.

Fáze technologického procesu zpracování obrazu

Proces zpracování obrazu je postupem vývoje obrazu o objektu, originálu v jednotlivých etapách pořízení a sběru dat, uložení a výběru dat, interpretace dat a implantace dat do originálu. Metodikou procesu zpracování obrazu (projektování informační soustavy) je kombinace postupů projektování „shora dolů“ a „zdola nahoru“ a kombinace induktivního a deduktivního přístupu v získávání poznatků o objektu jako pozorovaném prostředí. Základem je obecně používaná metodika schémat, založená na postupné transformaci uživatelských představ až do fungující formy informační soustavy. Základní tři úrovně této metodiky jsou:



malizované podoby dané podmínkami jazyka, forem datových prvků (datových modelů).

- Fyzické (vnitřní) schéma pojednává o transformaci logického schématu do prostředí, v němž bude informační soustava realizována, a ve kterém bude fungovat.

Kvalita obrazu, dat

Kvalita obrazu vypovídá o kvalitě dat zobrazující objekty (originály), jejich části a vlastnosti. Mezi nejvýznamnější soustavy typů, v nichž se vyskytují a používají data, patří data členěná podle úrovně konstrukce (bit, znak, slovo, věta, soubor, báze dat, báze znalostí) a podle úrovně významu ve smyslu pragmatické hodnoty.

Význam dat ve smyslu pragmatické hodnoty

Data členěná podle úrovně významu ve smyslu jejich pragmatické hodnoty jsou rozlišena v následujících úrovních:

- Data – relace poznatku k rozlišenému jazykovému konstruktu; jejich kvalita, spočívá v jejich zobrazení bez poznatkových souvislostí.
- Informace – relace údaje (dat) k poznatkovému úrovní uživatele. Jejím použitím se rozšiřuje dosud existující poznání uživatele.

- Znalost – relace údaje s kvalitou informace k jiným údajům (datům) nebo informacím. Počet dat či informací vstupujících do relace pak měří vyšší či nižší znalost.
- Odpovědnost – relace údaje s kvalitou informace nebo znalosti k efektu, výsledku aktivovaného procesu, vyvolává zvýšení kvality údaje v kvalitě odpovědnosti. Údaj nabyt kvality odpovědnosti, jestliže byl vstupem do procesu (aktivoval proces) vedoucí k nějakému výsledku, efektu.

Reprezentace znalostí

S rozlišením dat, obrazu podle kvality souvisí i rozlišení použití obrazu, uplatnění obrazu v chování objektu. V současnosti je vel-

ční pravidla, sémantické sítě, znalostní ontologie a další.

Produkční pravidla v expertních systémech

Systém produkčních pravidel je nejčastěji využíván v expertních systémech pro reprezentaci znalostí. Produkční systém je tvořen třemi částmi:

- Soubor produkčních pravidel, kde pravidla mají tvar situace → akce; levá strana pravidla (situační část pravidla) popisuje podmínku, za které může být pravidlo použito, pravá strana pravidla popisuje akci, která má být vykonána při splnění podmínky.
- Pracovní paměť obsahuje okamžitý stav řešené úlohy v podobě počátečních i odvozených dat.

kretizovány v podobě speciálnějších druhů, potomků, umožňuje vytvářet bohaté vnitřně členěné a přehledné komplexy znalostí s vysokou efektivitou odvozování nových znalostí a vyhledávání relevantních informací.

V sémantické síti lze díky taxonomickým vztahům množinové inkluze provádět odvozování specializací nebo generalizací. Při specializaci se informace v taxonomii přenášejí od obecnějších typů ke speciálnějšími, při generalizaci naopak. Takto odvozené objekty mohou přijímat, dědit vlastnosti od hierarchicky vyšších objektů, předků (odvozování pomocí specializace), resp. od hierarchicky nižších objektů, potomků (odvozování pomocí generalizace).

Sémantické sítě jsou základem dalších schémat pro reprezentaci znalostí. Jedná se například o rámce, relační datové modely nebo objektově orientované modely.

Znalostní ontologie

Podle definice T. Grubera je ontologie v teorii znalostního inženýrství chápána jako explicitní dohoda o sdílené konceptualizaci. Základní znaky ontologie v chápání znalostního inženýrství jsou:

- Ontologie je konceptuálním popisem znalostí – definuje co a v jaké podobě může být ve znalostech obsaženo. Její součástí jsou objekty a vztahy mezi objekty (ontologie nepopisuje konkrétní výskyty objektů).
- Ontologie je sdílitelná – daný způsob konceptualizace je přijat jako standard v rámci širší komunity. Člen komunity se zavazuje, že znalosti reprezentované danou ontologií (syntaxe a sémantika znalostí), jsou pro tohoto člena ontologickým závazkem (ontological commitment).
- Ontologie je definována explicitně – znalosti o prostředí jsou reprezentovány explicitně pomocí jazyka pro reprezentaci znalostí ve formě pravidel nebo poznatků (deklarativní nebo procedurální znalosti). Ontologie tak může být vyjádřena neformálně (volným textem) nebo formálně pomocí formálního jazyka.

Hlavním přínosem využívání ontologií je usnadnění komunikace mezi uživateli a pozorovateli objektu. Na konceptuální úrovni procesu zpracování obrazu o originálu se jedná o společný jazyk pro vyjádření poznatků o pozorovaném objektu. V rámci dalších úrovní může být ontologie základem pro společné rozhraní vzájemného porozumění mezi heterogenními objekty.

LITERATURA

- [1] Vlček, J., Inženýrská informatika. Vydavatelství ČVUT, Praha 1993.
- [2] Vlček, J., Systémové inženýrství. Vydavatelství ČVUT, Praha 1999.
- [3] Mařík, V., Štěpánková O., Lažanský J. a kol., Umělá inteligence (1). ACADEMIA, Praha 1999.
- [4] Melichar B., Jazyky a překlady. Vydavatelství ČVUT, 1999.
- [5] Moos P., Informační technologie. Nakladatelství ČVUT, Praha 1993.

ká pozornost věnována kvalitě dat v úrovni znalostí, jejich reprezentací a způsobů jejich využití.

Reprezentaci dat s kvalitou znalostí, které vstupují do funkcí prvků, procesů je možné klasifikovat v následujících úrovních:

- implicitní – znalosti jsou použité v podobě algoritmu funkce zachycené jazykem pro konstrukci algoritmu (např. programovací jazyk);
- explicitní – znalosti jsou oddělené od vlastní úlohy, funkce, která znalost zpracovává;
- deklarativní – znalosti svou reprezentací vyjadřují poznatky o množině objektů reálného světa, tzn. co je poznáno nebo dokázáno (např. pes je zvíře);
- procedurální – znalosti svou reprezentací vypočítávají o způsobu poznávání a odvozování; znalosti reprezentované procedurálně mají tvar pravidel (např. je-li X pes, potom je X zvíře).

Významnou roli v současném rozvoji informatiky představuje explicitní reprezentace znalostí v podobě pravidel nebo poznatků, umožňující jejich vícenásobné využití v rámci chování různých objektů, k nalezení výsledků, vysvětlení postupů jejich hledání, nebo odvozování nových znalostí. Takováto reprezentace znalostí je rovněž významná při řešení úloh vzájemného porozumění heterogenních částí systému.

Jazyk predikátové logiky 1. řádu je základním, výchozím nástrojem pro reprezentaci znalostí, který poskytuje prostředky pro popis objektů a relací mezi nimi. Na jeho základě jsou dále odvozovány jiné, příbuzné reprezentace, které zahrnují produk-

- Interpret pravidel (inferenční stroj) je řídicím algoritmem, který porovnává data v pracovní paměti s produkčními pravidly, vybírá pravidla vhodná k vykonání (provádí řešení konfliktů) a provede příslušnou akci.

Na obrázku na minulé straně je zobrazeno jednoduché schéma produkčního systému. Interpret pravidel může pracovat způsobem přímého zřetězení (forward chaining) nebo zpětného řetězení (backward chaining).

Sémantické sítě

Sémantické sítě umožňují zachytit znalosti – poznatky o objektech, jejich vlastnostech a relacích. V sémantických sítích je možné efektivně vyjadřovat hierarchické relace prostřednictvím vztahů mezi množinami (množinová inkluze, přináležitost k množině) a rozlišovat mezi typem objektu a instancí typu. Jednotlivé složky sémantické sítě jsou:

- typ objektu – obecný objekt reálného světa, oblast vymezená svými atributy a vztahy k jiným objektům (taxonomicky vyšším i nižším);
- instance objektu – konkrétní výskyt obecného typu objektu;
- množinová inkluze – relace vyjadřující vztah mezi objekty ve smyslu „je druhem“ (is-a, is-kind-of);
- přináležitost k množině – relace vyjadřující vztah mezi objekty ve smyslu „patří do“ (is-part-of).

V jedné sémantické síti je možné znalosti asociovat z různých hledisek. Sdružování objektů do společných obecnějších uskupení, tříd, které jsou na nižších úrovních kon-

