

Od sémantických sítí k logickým formalismům

Petr Křemen

FEL ČVUT

Co nás čeká

1 Vysvětlování modelovacích chyb

2 Black-box metody

- Algoritmy založené na CS-stromech
- Algoritmus založený na Reiterově algoritmu

Informace o předmětu

Informace o předmětu

- webová stránka:
<http://cw.felk.cvut.cz/doku.php/courses/a4m33rzn/start>
- čtyři témata: deskripční logika, temporální a modální logika, pravděpodobnostní modely, fuzzy logika
- pečlivě si projděte pravidla hry !

Co je to reprezentace znalostí ?

Motivace

- Mějme univerzitní doménu :
 - Student: **“Jaký předmět bakalářské etapy si mám zapsat, abych získal alespoň 6 kreditů ?”**
 - Učitel : **“Kolik hodin týdně budu tento semestr učit ?”**
 - Děkan : **“Které předměty jsou mezi studenty populární a které nikoliv ?”**
- Každý z nich potřebuje jiné informace - Co jsou tedy znalosti ?
- Znalosti se snaží postihnout vztahy a zákonitosti v dané doméně, tak aby mohly být využity k zodpovězení takovýchto dotazů.
 - **“Předměty bakalářské etapy jsou typem předmětů.”**
 - **“Ve většině případů je možné předmět otevřít pouze tehdy jsou-li zapsáni alespoň 2 studenti.”**
 - **“Je-li někdo vedoucím katedry, je též zaměstnancem školy.”**

Motivace (2)

Máme tedy doménu a znalosti. Ovšem:

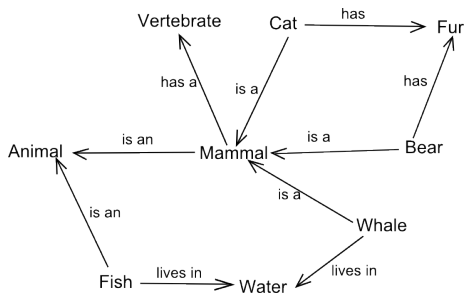
- jak tedy znalosti formálně reprezentovat ?
 - deklarativně × procedurálně ? - my nyní **deklarativně**. např.
 $(\forall P)(BakalarskyPredmet(P) \Rightarrow Predmet(P))$
 - bez neurčitosti (crisp) × s neurčitostí - my nyní **bez neurčitosti**. např.
 $(\forall K)(Kurz(K) \Rightarrow (KurzSVyjimkou(K) \vee ((\exists X_1, X_2)ZapsanNa(X_1, K) \wedge ZapsanNa(X_2, K) \wedge X_1 \neq X_2)))$
- jak využít výslednou reprezentaci ?
 - znalostní management - vyhledávače (databáze, sémantické servery, sémantický web)
 - multiagentní systémy - komunikační obsah zpráv zasílaných mezi agenty
 - strojové učení - jazykový bias
 - ... a vlastně všechny obory AI

Deklarativní reprezentace znalostí bez neurčitostí

- sémantické sítě, rámce,
- thesaury, mapy témat
- relační databáze (relační kalkulus)
- pravidlové systémy, Prolog (predikátová logika prvního řádu)
- sémantický web, RDF(S), OWL, OWL 2 (deskripční logiky)

Sémantické sítě

Sémantické sítě



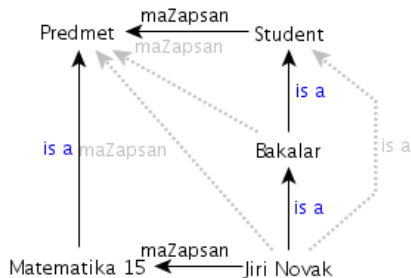
(©wikipedia.org)

- Uzly jsou tvořeny entitami (instance, třídy), hrany reprezentují binární relace.
- Jediná inference je dědění pomocí **is a** relace.

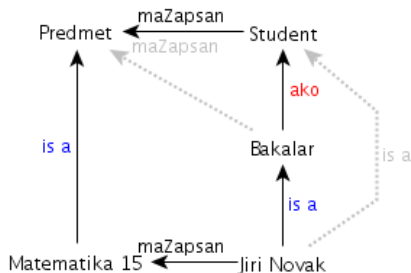
Příklad

Each Cat **has a** Vertebrate, since

Sémantické sítě (2)



Tento způsob však nerozlišuje jednotlivce (instance) a množiny (třídy) ...



Řešením je zavést nový typ relace "is a kind of" **ako** a používat ji pro dědičnost, zatímco **is a** používat pro instanciování.

Sémantické sítě (3)

- 😊 jsou velmi jednoduché - z logického pohledu se jedná o jednoduchou binární relační strukturu – vyjma relací **ako** a **is a**. Sémantiku těchto relací lze vyjádřit univerzálními uzávěry formulí

$$relace(X, Y) \wedge ako_a(Z, X) \Rightarrow relace(Z, Y).$$

$$isa(X, Y) \wedge ako(Y, Z) \Rightarrow isa(X, Z).$$

$$ako(X, Y) \wedge ako(Y, Z) \Rightarrow ako(X, Z).$$

- ☹ neumožňují vyjádřit nemonotónní znalosti (podobně jako FOL).
- ☹ neumožňují vyjádřit n-ární relace. Ty je nutné nejprve **reifikovat**.
- ☹ neumožňují vyjádřit vlastnosti binárních relací - tranzitivita, funkcionalita, reflexivita, atd, ani jejich hierarchie “býti otcem znamená i býti rodičem”, aj.,
- ☹ neumožňují vyjádřit složitější konstrukty, jako kardinality: “Každý člověk má nejvýše dvě nohy.”
- Wordnet, sémantické wiki, aj.

Sémantické sítě – Wordnet, MultiWordnet

Wordnet (<http://wordnet.princeton.edu>) a MultiWordnet (<http://multiwordnet.itc.it>) jsou lexikální databáze. Jedná se o sémantické sítě, které rozšiřují dosud jmenované relace o další sémantiku, např. :

hyponyma, hypernyma odpovídají **ako** relaci.

meronyma, holonyma označují “part-of” vztahy mezi pojmy.

synonyma, antonyma synonyma se sdružují v tzv. *synsetech* - jedná se o množiny termů, odpovídající jednomu sémantickému kontextu (např.

$$S_1 = \{\text{man, adult male}\}, S_2 = \{\text{man, human being}\}$$

Rámce

Rámce

frame: Škoda Favorit

slots:

is a: osobní auto

ma motor: čtyřdobý benzínový

ma převodovku: manuální

ma karburator: *hodnota:* Jikov

předpoklad: Pierburg

- strukturovanější než SN
- formuláře, které obsahují **slots** (binární relace).

([MvL93])

- Každý slot může mít několik **facetů** (omezení na používání slotu), např. kardinalitu, defaultní hodnotu, apod.
- 😊 Facety umožňují nemonotónní odvozování.
- 😊 Lze definovat *démony* – triggerery pro akce prováděné na facetech (čtení, změna, smazání). Lze je použít např. pro ověřování konzistence.

Rámce (2)

Příklad

Škoda Favorit typicky **má karburátor** typu Pierburg, tento Škoda Favorit však **má karburátor** typu Jikov.

- rámce lze sdružovat do *scénářů*. Ty představují typické situace, např. návštěva restaurace, apod. [MvL93]
- OKBC - <http://www.ai.sri.com/okbc>
- Protégé - <http://protege.stanford.edu/overview/protege-frames.html>
- Apollo - <http://apollo.open.ac.uk>
- Apollo CH - <http://labe.felk.cvut.cz/falc/Apollo>

Rámce (3) - Apollo CH

The screenshot shows the Apollo 0.28.0 application window. The interface is divided into several panes:

- Left Pane (Class Hierarchy):** Shows a tree structure starting with 'default', containing 'time-and-date', 'diagnose-and-fix', and 'general-medical-knowledge'. Under 'general-medical-knowledge', there are sub-classes like 'drug', 'my-class', 'blood-pressure', 'disease-disorder', 'cerebrovascular-disease', 'heart-disease', 'renal-disease', 'vascular-disease', 'medical-condition-disorder', and 'ace-inhibitors'.
- Top Pane (Super-classes):** Lists 'disorder' as the super-class for the selected class.
- Right Pane (Slots using focus):** Currently empty.
- Bottom Pane (Property Details):** Shows details for the 'description' slot of the 'disease-disorder' class.

Slot	Type	Value	Cardinality	Documen.
description	string		R	
- Bottom Right Pane (Sub-classes and instances):** Lists 'cerebrovascular-disease', 'heart-disease', 'renal-disease', and 'vascular-disease' as sub-classes.

Sub-classes and instances	Classes used by focus
cerebrovascular-disease	string
heart-disease	
renal-disease	
vascular-disease	

The status bar at the bottom indicates the current class is 'general-medical-knowledge'.

Rámce (4) - Protégé

The screenshot displays the Protégé 3.2.1 interface. The main window title is "newspaper Protégé 3.2.1 (file:/home/kremen/programs/Protege_3.2.1/examples/newspaper/newspaper.pprj, Protégé Files (.pont and .pins))". The menu bar includes File, Edit, Project, Window, Tools, and Help. The toolbar contains various icons for file operations and editing.

The interface is divided into several panes:

- Class Browser:** Shows a hierarchy for the project "newspaper". The selected class is ".STANDARD-SLOT".
- Class Editor:** Shows the details for the selected class ".STANDARD-SLOT" (instance of :STANDARD-CLASS). It includes fields for Name, Documentation, and Constraints. The "Rule" section shows "Concrete" selected.
- Template Slots Table:** A table listing various slots and their properties.

Name	Cardinality	Type	Other Facets
:ASSOCIATED-FACET	single	Instance of FACET	inverse-slot=ASSOCIATED-SLOT
:DIRECT-DOMAIN	multiple	Instance of CLASS	inverse-slot=DIRECT-TEMPLATE-SLOTS
:DIRECT-SUBSLOTS	multiple	Instance of SLOT	inverse-slot=DIRECT-SUBSLOTS
:DIRECT-SUPERLOTS	multiple	Instance of SLOT	inverse-slot=DIRECT-SUBSLOTS
:DIRECT-TYPE	multiple	Class with superclass :SLOT	inverse-slot=DIRECT-INSTANCES
:DOCUMENTATION	multiple	String	
:NAME	single	String	
:SLOT-CONSTRAINTS	multiple	Instance of :CONSTRAINT	
:SLOT-DEFAULTS	multiple	Any	
:SLOT-INVERSE	single	Instance of SLOT	inverse-slot=SLOT-INVERSE default=1
:SLOT-MAXIMUM-CARDINALITY	single	Integer	
:SLOT-MINIMUM-CARDINALITY	single	Integer	
:SLOT-NUMERIC-MAXIMUM	single	Float	
:SLOT-NUMERIC-MINIMUM	single	Float	
:SLOT-VALUE-TYPE	multiple	Any	default=String
:SLOT-VALUES	multiple	Any	

Rámce a sémantické sítě - shrnutí

- 😊 velmi jednoduché struktury pro reprezentaci znalostí,
- 😊 nemonotónní odvozování,
- 😞 ad-hoc odvozovací procedury, překlad do FOL není jednoduchý, a tedy ani jednoznačný,
- 😞 problémy – dotazovací jazyk, debugging.
 - na sémantických sítích staví dnešní :
 - **thesaury**
 - **mapy témat**

Thesauri

Thesaurus

taxonomie je hierarchie pojmů

thesaurus je taxonomie obohacená o další typy relací. Může být buď jednojazykový (ISO 2788:1986) nebo vícejazykový (ISO 5964:1985). Příklady použitých relací :

BT/NT (broader/narrower term) = hierarchie pojmů.

Příklad

skopové maso → BT → maso

SN (scope note) vysvětluje význam daného termu.

Příklad

škola → SN → instituce sloužící k vzdělávání

USE/UF slouží k popisu deprecated synonym

Příklad

vteřina → USE → sekunda.

Thesaurus – pro a proti

- 😊 čitelnější než striktně formální jazyky pro sémantický web – snadněji tvořitelné.
- ☹ opět problémy se sémantikou :

Příklad

Relace BT může být použita ve významech:

subsumpce , např. jablko BT ovoce,

instance , např. David BT člověk,

být částí , např. kapota BT auto.

...

Mapy témat

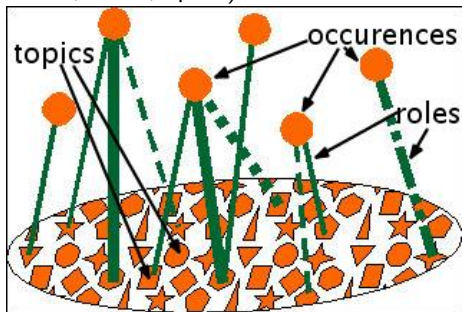
Mapy témat – témata

- ISO standard – ISO/IEC 13250:2003
- tři typy objektů : **témata** (topics), **výskyty** (occurrences) a **asociace** (associations).
- témata
 - představují konceptuální pojmy - třídy, instance, vlastnosti, atd.
 - téma může mít několik tzv. **typů témat** (topic types). Vztah “mítí typ” vytváří hierarchii témat (analogie *isa* relace v sémantických sítích, nebo vlastnosti *rdf:type* v RDF(S), viz. příští přednáška).
 - tématu může být přiřazeno několik **jmen** (např. přezdívka, formální jméno, přihlašovací jméno, atd.), každé z nich v několika **variantách** (např. zobrazení vs. třídění).

Mapy témat – výskyty

- výskyty

- představují “odkazy” témat na reálné dokumenty/informační zdroje.
- téma je s výskytem svázáno pomocí tzv. **role**, která určuje typ výskytu (web. stránka, článek, kniha, apod.).

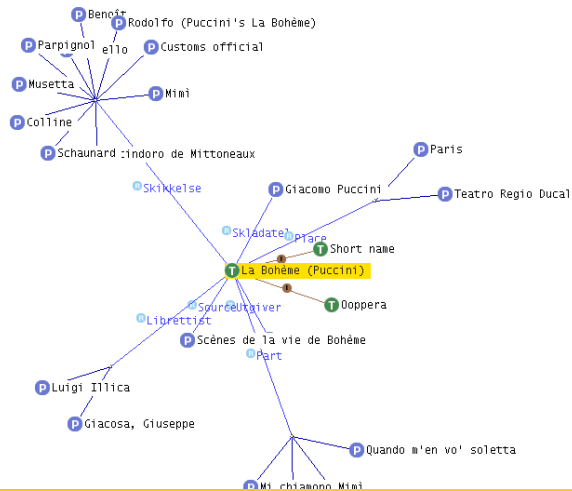


(<http://www.ontopia.net/topicmaps/materials/tao.html>)

Mapy témat – asociace

- asociace
 - představují vztahy mezi tématy – analogie n-árních relací
 - asociaci je přiřazen **asociační typ**, který je tématem (,a dále typ tématu je speciálním asociačním typem).
 - témata vystupují v asociacích v tzv. **asociačních rolích**
 - též asociační roli je přiřazen **typ asociační role**, který je tématem

Mapy témat – příklad

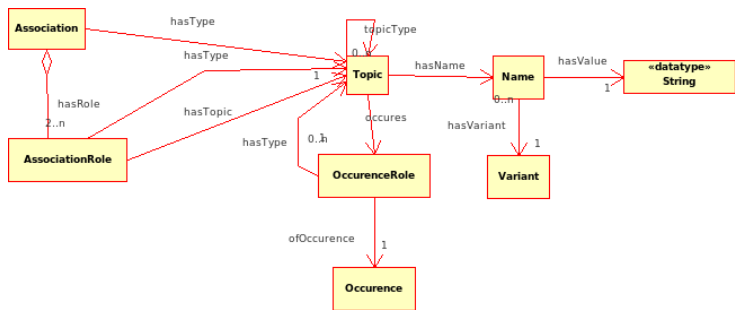


T ... témata

P ... částečně
rozvinutá témata
(kromě typů
témat)

R ... asociace

Mapy témat – model



Nástroje a odkazy

- nástroje:
 - Ontopia (Ontopoly, Omnigator, Vizigator) – hlavní tahoun vývoje v TM
 - TM4L
 - TM4J
 - ... a mnoho dalších
- odkazy:
 - <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
 - <http://www.kosek.cz/xml/tmtut/>

TM4L Viewer

The screenshot displays the TM4L Viewer interface with three main panes:

- Topic Maps:** Shows the file path: `M:\file:\home\kremen\downloads\zsyopera.xtm`.
- Tree View:** A hierarchical tree structure for 'La Bohème (Puccini)'. The root node is 'La Bohème (Puccini)', which branches into 'Overture', 'Short name', 'Web page', 'Skikkelse', 'Place', 'Libretto', 'Sound clip', 'Article', 'Libretto', 'Illustration', 'Synopsis', 'Utgiver', 'Ricordi', 'Première date', 'Part', and 'Log'. Further sub-nodes include 'Paris', 'Teatro Regio Duca', 'Giuseppe Puccini', 'Scènes de la vie de Bohème', 'Giacosa, Giuseppe', 'Luigi Illica', 'Puccini', 'Ricordi', 'Short name', 'Skikkelse', 'Part', 'Customs official', 'Marcello', 'Mimi', 'Musetta', 'Parpignol', 'Rodolfo (Puccini's La Bohème)', 'Schaunard', 'Benoît', 'Che gelida manina', 'Mi chiamano Mimi', and 'Quando m'en vo' sole'.
- Graph/Text View:** A network graph where nodes represent the entities from the tree view, connected by lines representing relationships. The central node is 'La Bohème (Puccini)'. Other nodes include 'Paris', 'Teatro Regio Duca', 'Giuseppe Puccini', 'Scènes de la vie de Bohème', 'Giacosa, Giuseppe', 'Luigi Illica', 'Puccini', 'Ricordi', 'Short name', 'Skikkelse', 'Part', 'Customs official', 'Marcello', 'Mimi', 'Musetta', 'Parpignol', 'Rodolfo (Puccini's La Bohème)', 'Schaunard', 'Benoît', 'Che gelida manina', 'Mi chiamano Mimi', and 'Quando m'en vo' sole'.
- Log:** A message at the bottom states: 'Cannot activate TopicMapReference, because its in state StoreState (opening)'.

http://localhost:8080/omniator/model/topic_complete.jsp?term=zyopera.stm&id=la-boheme

omniator Home | Manage | Wikidata | Support | About

zyOpera | Customiz | Filter | Export | Merge | Statistics | Query | Edit | No schema | Vypiste

La Bohème (Puccini) Type(s): Opera

Untyped Names (1)

- **La Bohème (Puccini)**
 - Bohème (Puccini) - Scope: Sort

Short name (1)

- **La Bohème** - Scope: Puccini, Giacomo
 - Bohème - Scope: Puccini, Giacomo: Sort

Associations (21)

- **Based on**
 - Scènes de la vie de Bohème
- **Composed by**
 - Puccini, Giacomo
- **Cast/alm**
 - Che gelida manina
 - Mi chiamano Mimì
 - O soave fanciulla
 - Quando m'en va' soletta
- **Dramatic personas**
 - Alcindoro de Mitzouneaux
 - Benoit
 - Colline
 - Customs official
 - Marcello
 - Mimì
 - Musetta
 - Parginal
 - Rodolfo (Puccini's La Bohème)
 - Schaunard
- **First performed at**
 - Teatro Regio Duca
- **Libretto by**
 - Giacosa, Giuseppe
 - Illica, Luigi
- **Published by**
 - Ricordi
- **Takes place in**
 - Paris

Subject Identifiers (1)

- http://en.wikipedia.org/wiki/La_Bohème

Internal Occurrences (2)

- **Première date**
 - 1896-02-01
- **Video recording**
 - 100 046

External Occurrences (11)

- **Article**
 - http://en.wikipedia.org/wiki/La_Bohème - Scope: Web: Wikipedia
- **Illustration**
 - <http://localhost:8080/operamap/oc/curs/operapuccini/la-boheme-poster1.jpg> - Scope: Local
- **Librette**
 - <http://opera.stanford.edu/operapuccini/LaBoheme/1bretto.html> - Scope: Opera Glass: Web
- **Poster**
 - <http://www1-ds.com/operapuccinian/pictures/boheme-score.htm> - Scope: OperaResource: Web
 - http://www1-ds.com/operapuccinian/pictures/bohema_poster.htm - Scope: OperaResource: Web
- **Sound clip**
 - <http://localhost:8080/operamap/oc/curs/la-boheme.wav> - Scope: Local
 - <http://www.ontopia.net/topicmaps/examples/operapuccini/la-boheme.wav> - Scope: Web
- **Synopsis**
 - http://localhost:8080/operamap/oc/curs/la_syn.php3.htm - Scope: Arizona Opera: Local
 - <http://www.azopera.com/learn/synopsis/boheme.shtml> - Scope: Arizona Opera: Web
 - <http://www.metopera.org/synopses/boheme.html> - Scope: Opera News: Web
- **Web page**
 - <http://opera.stanford.edu/operapuccini/LaBoheme/main.html> - Scope: Opera Glass: Web

Scoped Names (1)

- **Rodolfo (Puccini's La Bohème)** (Rodolfo (Puccini's La Bohème))

Scoped Association Types (2)

Další aspekty map témat

- kromě přehledu uvedeného výše je možné sdružovat mapy témat (spec. typ tématu) do **kontextů** (scopes, themes). Kontexty představují filtrační mechanismus

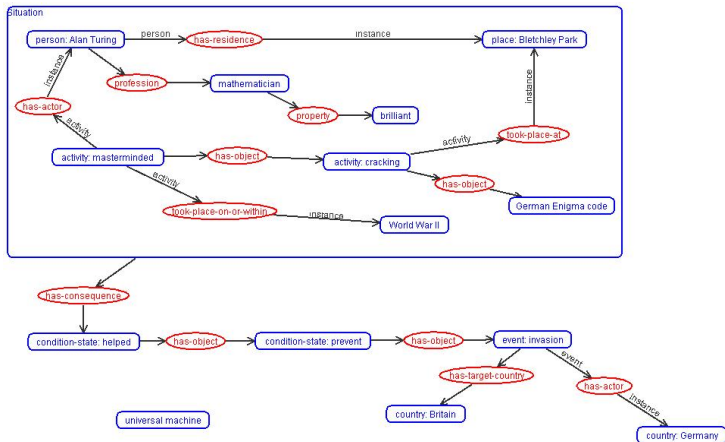
Příklad

Modelujeme-li doménu *turismus*, jiná data by měla být dostupná zájemci o cestování a jiná pracovníkovi cestovní kanceláře.

- XTM je XML formát pro ukládání map témat.
- dotazování pomocí TMQL, nebo tologu (podobná syntax jako SQL).

Konceptuální grafy

Příklad



Konceptuální grafy

konceptuální graf je bipartitní graf s uzly typu (1) **koncept** a (2) **relace**.
koncept je tvaru **typ-konceptu : referent**.

Příklad (Typy kvantifikátorů)

dog : Lucky

“Pes Lucky”

$\exists x \text{Pes}(x) \wedge \text{Name}(x, \text{Lucky}) \wedge \dots$

dog

“Nějaký pes”

$\exists x \text{Pes}(x) \wedge \dots$

dog : \forall

“Všichni psi”

$\forall x \text{Pes}(x) \rightarrow \dots$

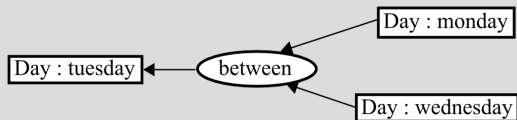
dog : {*}

“Množina psů”

☹ není FOL

konceptuální relace = vztah = predikát libovolné arity > 0 .

Příklad (ternární relace)

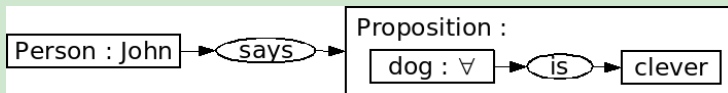


Konceptuální grafy (2)

referent se skládá (viz. předchozí slide) z **kvantifikátoru** (existenciální, nebo definovaný (univerzální, kolektivní, aj.)), **designátoru** (ten označuje identifikátor instance, např. jméno) a případně tzv. **deskriptoru** (konceptuální graf popisující daný koncept).

kontext je koncept s neprázdným deskriptorem

Příklad (Kontext)



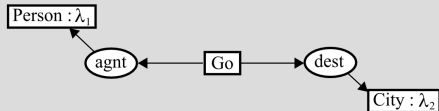
“John říká, že všichni psi jsou chytrí.”

☹ není FOL

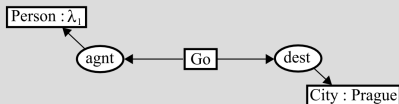
Konceptuální grafy (3)

lambda výrazy představují jakási “makra” – umožňují definovat konceptuální relace pomocí “vzoru” konceptuálního grafu. “Dosazované” proměnné se označí symboly λ_i .

Příklad (lambda výrazy)



def. binární relaci “Go”.

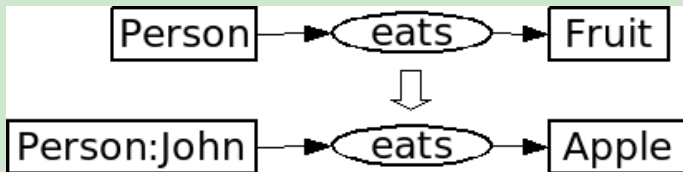


def. unární relaci “Go to Prague”.

Konceptuálních grafech – inference

- inference využívá několika forward chaining pravidel¹ (zobecnění grafu, specializace grafu, ekvivalentní úpravy).
- vyhledávání (jakožto složitější inferenční procedura) se provádí pomocí tzv. **projekce**. Ta hledá výskyt vzoru konceptuálního grafu v grafu odpovídajícím dané znalostní bázi s využitím hierarchie konceptuálních typů a typů konceptuálních relací.

Příklad (projekce)



¹<http://www.jfsowa.com/cg/cgstandw.htm>

Konceptuální grafy – nástroje

CharGer – editor CG

(<http://sourceforge.net/projects/charger>)

Notio – Java knihovna + API pro manipulaci s CG

(<http://backtrack.uwaterloo.ca/CG/projects/notio>)

Prolog+CG – inferenční stroj pro CG v Prologu

(<http://prologpluscg.sourceforge.net>)

Amine – novější verze Prolog+CG

(<http://amine-platform.sourceforge.net>)

DNA – anotační nástroj využívající pro vizualizaci CG

(<http://labe.felk.cvut.cz/~uhlir/DNATWeb/DNATHome.html>)

Amine4

The screenshot displays the Amine4 software interface, titled "CG Operations Interface - Untitled". It features a menu bar (File, Edit, Font, Parameters, Ontology, CG Operations, Animation, Help) and a toolbar. The main workspace is divided into three panes, each showing a conceptual graph (CG) in English and French.

- Input CG1:** Shows a simple graph with a node "Sex = male" connected to "Human:super".
- Input CG2:** Shows a more complex graph with nodes "Object", "Push", and "Human". Edges include "obj", "sgrt", and "on".
- Output CG:** Shows a graph with nodes "Sex = male", "Human", "Push", and "Object". Edges include "obj", "sgrt", "on", and "rely".

On the left side, there is a sidebar with various operations: Equal, MaximalJoin, Generalize, Subsume, SubsumeSk, Analogy, Compare, CoveredBy, IsCanonic, Expand, and Contract. At the bottom, the status bar indicates the current lexicon: "Current Lexicon:Ontology : /home/kremen/programs/Amine4/samples/ontology/ManOntologyContext.xml".

- editace/prohlížení ontologií
- editace/prohlížení CG
- operace nad CG – příklad: JOIN
- inference CG+Prolog
- multiagentní systémy

Konceptuální grafy – shrnutí

- CG's (J.F. Sowa 80's) jsou představitelem formálních (strojově zpracovatelných) a přitom dobře čitelných, intuitivních jazyků,
- vycházejí myšlenkově z Pierceho existenciálních grafů [Sow00], [Dau01],
- jsou expresivnější než logika prvního řádu – nerozhodnutelnost,
- předchozí problém řeší tzv. *simple graphs* (J.F. Sowa 80's), které omezují tvar referentu a neumožňují tvořit kontexty.

Přehled – a co dál ?

- nyní jsme pouze přehledově prošli některé důležité milníky ve vývoji moderní reprezentace znalostí.
- ☹ tyto přístupy většinou mají problémy s formální sémantikou – nutnou podmínkou pro automatické zpracování rozsáhlých souborů informace.
- my se nyní podíváme na jazyky, které tuto formální sémantiku nepostrádají a přesto mají v jistém smyslu dobré výpočetní vlastnosti