



Laboratory for
information systems

Rudjer Boskovic Institute, Croatia

Ontologije u znanstvenim primjenama

Dragan Gamberger

Institut Rudjer Bošković

<http://www.irb.hr/home/gamber/>

17.11.2010



Pregled izlaganja

- Laboratorij za informacijske sustave
- Europski projekti na kojima radimo i potreba za ontologija
- Osnovni koncepti i značaj razvoja ontologija
- Naša iskustva sa primjenom ontologija
 - medicinska domena zatajenja srca
 - data mining ontologija
- Raskorak između velikih očekivanja i skromne primjene kao poticaj istraživanjima

Dio Zavoda za elektroniku

Predmet istraživanja: tehnologije znanja

3 projekta unutar programa: RAČUNARSKI
POSTUPCI OTKRIVANJA ZNANJA U
ZNANSTVENIM PRIMJENAMA

- 3 znanstvenika
- 1 tehničar
- 3 zn. novaka
- 2 istraživača na EU projektima



projekti:

Algoritmi strojnog učenja i njihova primjena

Strojno učenje prediktivnih modela u računalnoj biologiji

Postupci računalne inteligencije u mjernim sustavima

Istraživanje u području:

tehnologije znanja - (područje: umjetna inteligencija - računarstvo)

strojno učenje

- inteligentna analiza podataka
- otkrivanje znanja
- podacima intenzivni procesi

sistematizacija znanja i primjene

- prikaz znanja
- potpora odlučivanju

znanstvene primjene

- tehničke znanostima
- biologija
- medicina
- kemija
- društvene znanosti
- ekonomija

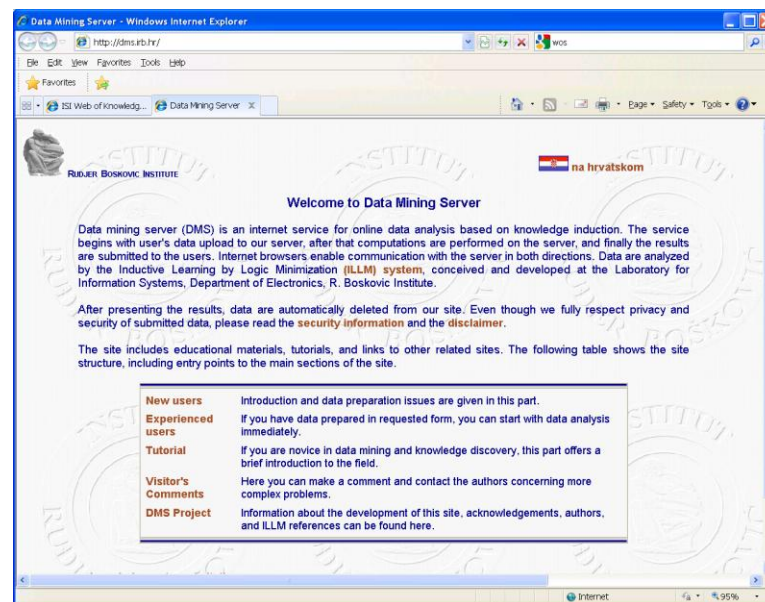
Dugogodišnji rad na razvoju postupaka strojnog učenja i njihovim primjenama

ILLM – vlastiti sustav strojnog učenja razvijen 1990-2000 koji uključuje

- učenje pravila
- otkrivanje podgrupa
- detektiranje i izdvajanje relevantnih značajki
- otkrivanje izuzetaka na osnovi saturacije skupa za učenje

DMS - mrežna usluga za otkrivanje znanja koja omogućuje da se vlastiti postupci mogu javno koristiti

- aktivna od 2000. godine
- jedna od globalno prvih mrežnih usluga ove vrste
- prosječno 5 primjena na dan



Bitno određuju aktivnosti unutar laboratorija

FP6 **HEARTFAID:**

2006-2009

STREP

“A knowledge based platform of services for supporting medical-clinical management of heart failure within elderly population”

Vodili smo važan dio:

Knowledge representation, discovery and management

Financijaki dio za IRB: 200.000 EUR,

5 mladih istraživača koji su radili 6 mjeseci - 3 godine

nova oprema

međunarodna prepoznatljivost, niz novih prijedloga projekata, zajedničke primjene ..





T4.4 - Ontologies and medical knowledge representation in the domain

... presentation of knowledge collected from medical experts and gathered during data analysis in a form appropriate for diagnosis explanation and decision making processes. ..

T5.4 – Implementation of the decision support system

... designing and developing the prototype of the “intelligent” decision support system. This engine will be based on a knowledge base specifically developed to model the knowledge of the domain experts. ...

T4.4 - Ontologies and medical knowledge representation in the domain

... presentation of knowledge collected from medical experts and gathered during data analysis **in a form appropriate for** diagnosis explanation and decision making processes.

T5.4 – Implementation of the decision support system

... designing and developing the prototype of the “intelligent” decision support system. This engine will be based on a knowledge base specifically developed to model the knowledge of the domain experts. ...

This is a core **activity of the entire project** since it is aimed at designing the inference-reasoning engine of the platform that will provide the clinicians with an automatic support to the diagnostic practice.



FP7 **eLICO**:
2010-2012
STREP

“e-Laboratory for Interdisciplinary Collaborative Research in Data Mining and Data-Intensive Sciences”

naši zadaci:

- razvoj novih složenih procesa za analizu podataka (sustavi preporučavanja usluga)
- implementacija novih funkcionalnosti u procesima otkrivanja znanja
- sudjelovanje u razvoju “Data Mining Ontology” i njenoj primjeni na optimizaciju procesa

Financijski dio IRB: 150.000 EUR,

2 mlada istraživača

novi, međunarodno vrlo kompetentan konzorcij partnera (Švicarska, Velika Britanija, Njemačka)



Abstract:

... a **knowledge-driven** data mining assistant will rely on a **data mining ontology** and **knowledge base** to plan the mining process and propose ranked workflows for a given application problem.

O5.1

To build and maintain a task **ontology** on structured data mining that will serve as reference for professionals and users of data mining technology, **independently of the application domain**.

WP7.7

Develop **ontology-based approaches to meta-learning**. Design, test, and evaluate **ontology-based ranking methods** for providing a ranked list of workflows ..



Ontologije

"a concise and unambiguous description of what principle entities are relevant to an application domain and the relationships between them" (Schulze-Kremer, 2001)

Glavne reference:

Open Biomedical Ontologies (OBO) Foundry

<http://www.obofoundry.org/>

rezpozitorij korisnih ontologija za biomedicinu
specificira dobru praksu razvoja ontologija

Basic Formal Ontology (BFO)

<http://www.infomis.org/bfo>

ontologija visoke razine
općenit okvir za prikaz znanstvenih istraživanja

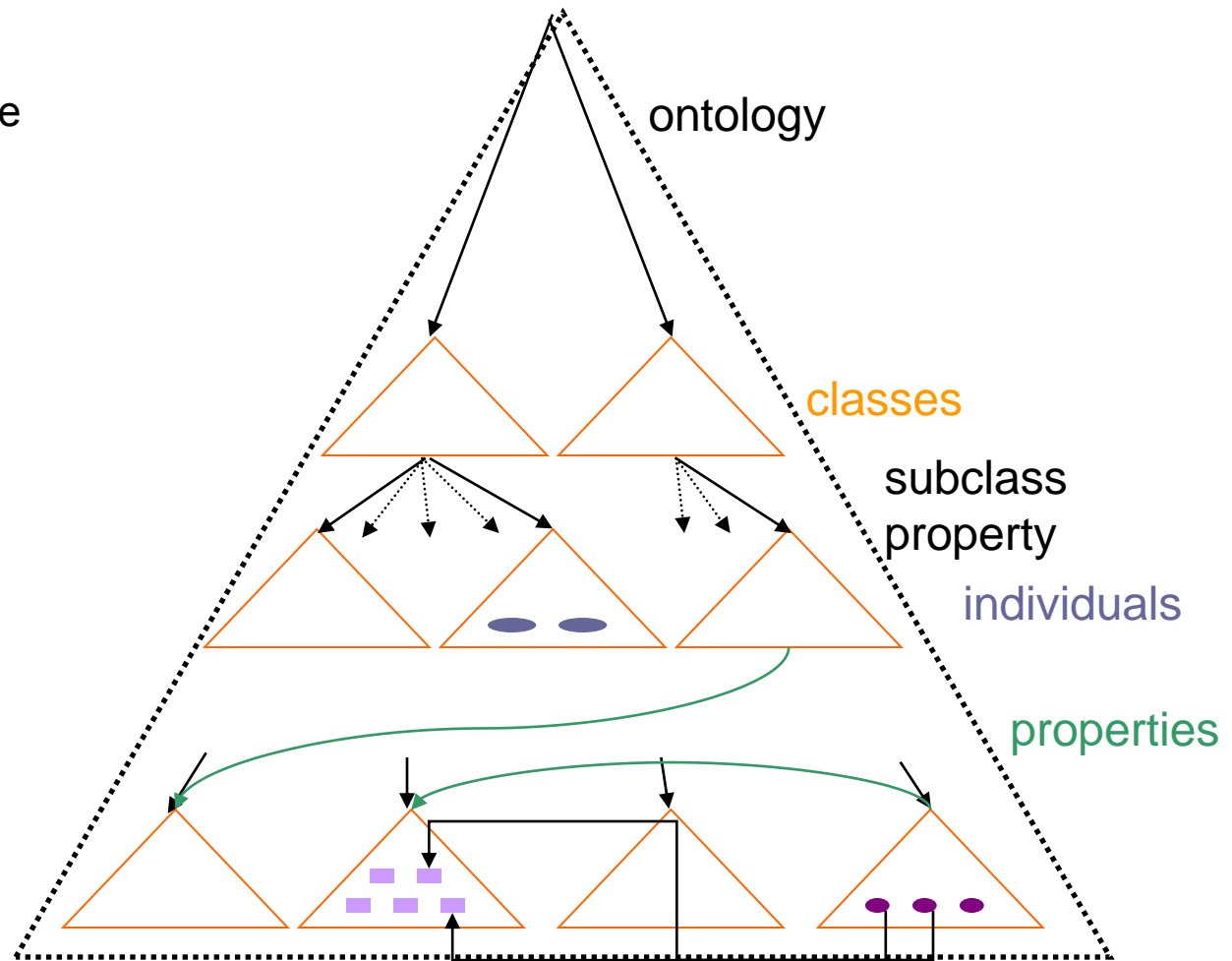
Ontologije

Ontologije su široko prihvaćene kao prikladna forma za konceptualizaciju znanja.

a) rječnik područja (**terminologija**),




b) organizacija pojmova u hijerarhijskoj strukturi (**taksonomija**),

c) definicija relacija između koncepata (**ontologija**).





Očekivanja od razvoja ontologija

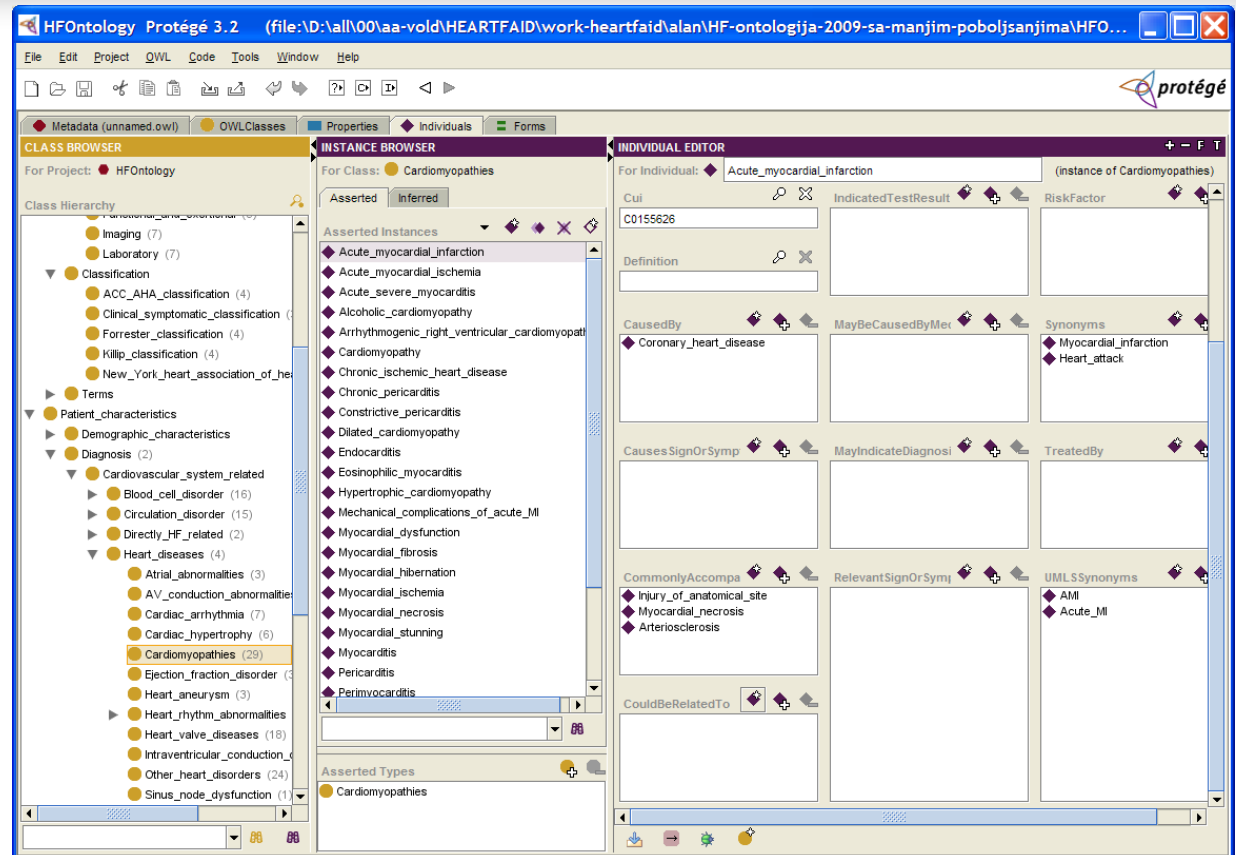
- a) - omogućavaju jednostavniju implementaciju i održavanje složenih sustava za obradu podataka i znanja (ekspertni sustavi, analiza podataka, planiranje). 
- b) - mogu se koristiti u raznim primjenama, osiguravaju njihovu interoperabilnost, verifikaciju njihove korektnosti te usporedivost rezultata raznih primjena. 
- c) - korišteni pojmovi moraju biti precizno definirani što je važno za tehničke sustave ali i kao osnova međuljudske komunikacije uključujući znanstvene radove, udbenike i stručne procedure (guidelines). 

Alat: Protégé

Ontološki jezik: OWL
(Web Ontology Language)

Logički formalizam:
deskriptivna logika

znajčajno utječe na
ekspresivnost
prikaza znanja ali
istovremeno i na
mogućnost
zaključivanja





Današnji izazovi

- Razvoj alata:** Protégé 3.2 -> Protégé 4.1
- Zaključivanje:** Algoritmi zaključivanja za razne ekspresivnosti deskriptivne logike (Jena, Pellet, Tableau algoritam)
osnovni teoretski problem i predmet istraživanja !
- Sakupljanje znanja:** Ujednačeni načini prikaza znanja, baze ontologija, interoperabilnost (Open Biomedical Ontologies (OBO), Gene Ontology Consortium).
- Primjene:** direktna integracija u ekspertne sustave, sustave za optimizaciju i planiranje, korištenje ontološkog znanja u *procesu otkrivanja novog znanja* iz podataka.



Primjeri ontologija na kojima radimo

Ontologija zatajenja srca

Data Mining ontologija (u tijeku)

Mali ekspertni sustav zasnovan na integraciji deklarativnog i proceduralnog znanja

Zatajenje srca

ontologijaGVumisD5OWL2 Protégé 3.2 (file:\D:\all\00\aa-vold\HEARTFAID\work-heartfaid\alan\HF-ontologija-2008-new-guideleines\...)

File Edit Project OWL Code Tools Window Help

protégé

Metadata (unnamed.owl) OWLClasses Properties Individuals Forms

CLASS BROWSER
For Project: ontologijaGVumisD5OWL2

Class Hierarchy

- swrl:Builtin (78)
- swrl:Imp (1)
- swrl:Variable (1)
- HF_concept
 - CHF_risks
 - Classification
 - Terms
 - Timestamp
- Patient_characteristics
 - Demographic_characteristics
 - Diagnosis (2)
 - Other_patient_characteristics (1)
 - Prognosis (5)
 - Signs_and_symptoms (4)
- Patients (3)
- Testing
 - Normal_ranges (33)
 - Physical_examination (1)
 - Relevant_test_results (107)
 - Test_characteristics (3)
 - Test_list
 - Test_measurements (2)
- Treatment (3)
 - Additional_therapy (21)
 - Medical_devices_and_surgical_procedure
 - Medical_procedure (13)
 - Medication
 - Patient_education (12)
 - Recommendations (15)

INSTANCE BROWSER
For Class: Medical_procedure

Asserted Inferred

Asserted Instances

- Cardiac_auscultation
- Carotid_endarterectomy
- Continuous_positive_airway_pressure
- Electric_countershock
- Endotracheal_intubation
- Infusion
- Inhalation_anesthesia
- Non-invasive_ventilation
- Non_invasive_ventilation_with_positive_end-expirato
- Oral_rehydration
- Spinal_anesthesia
- Thoracentesis
- Thrombolysis

Asserted Types

- Medical_procedure

INDIVIDUAL EDITOR
For Individual: Carotid_endarterectomy (instance of Medical_procedure)

Property	Value	Lang
rdfs:comment		

Cui: C0014099

Definition: nica intima of a carotid artery. (MeSH)

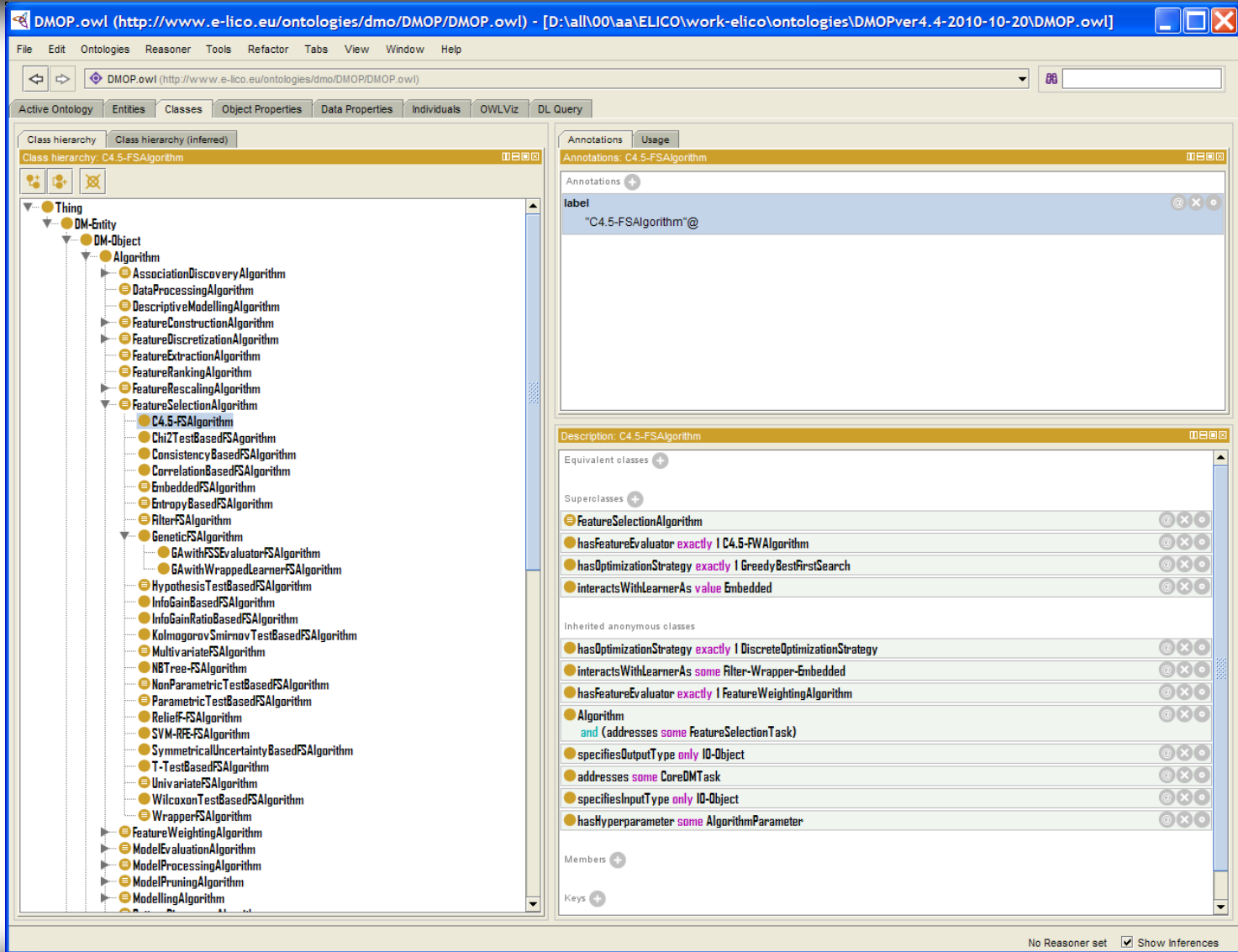
UMLSynonyms: Carotid_thromboendarterectomy



Zatajenje srca <http://lis.irb.hr/heartfaid/ontology/>

The screenshot shows a Windows Internet Explorer browser window. The title bar reads "HF Ontology in accordance with guidelines of European Society of Cardiology published on 30 Aug - Windows Internet Explorer". The address bar contains the URL "http://lis.irb.hr/heartfaid/ontology/". The browser's Favorites bar is visible, showing several bookmarks including "Heartfaid KB", "GlassFish Serve...", "obo ontology -...", "Open Biomedic...", and "HF Ontolog...". The main content area displays the "protégé" logo and the title "HF Ontology in accordance with guidelines of European Society of Cardiology published on 30 August 2008 (last update on September 14, 2009)". Below the title is a hierarchical tree view of the ontology classes, including:

- HF_concept
 - CHF_risks
 - Clinical (8 instances)
 - Demographics (9 instances)
 - Electrophysiologic (8 instances)
 - Functional_and_exertional (5 instances)
 - Imaging (7 instances)
 - Laboratory (7 instances)
 - Classification
 - ACC_AHA_classification (4 instances)
 - Clinical_symptomatic_classification (3 instances)
 - Forrester_classification (4 instances)
 - Killip_classification (4 instances)
 - New_York_heart_association_of_heart_failure_symptoms (5 instances)
 - Terms
 - Epidemiology_terms (4 instances)
 - Synonym (81 instances)
 - UMLS_syn (694 instances)
 - Treatment (3 instances)
 - Additional_therapy (21 instances)
 - Medical_devices_and_surgical_procedures
 - Medical_device (12 instances)
 - Artificial_cardiac_pacemaker (3 instances)
 - Cardiac_resynchronization_therapy (2 instances)
 - Surgical_procedure (5 instances)
 - Heart_valve_surgery (4 instances)
 - Left_ventricular_restoration (4 instances)
 - Revascularization (3 instances)
 - Medical_procedure (13 instances)
 - Medication
 - Avoid_or_use_with_caution_medications (9 instances)
 - Heart_failure_medication_group (20 instances)
 - ACE_inhibitor (5 instances)
 - Adrenergic_beta_antagonist (8 instances)
 - Aldosterone_receptor_antagonist (2 instances)
 - Angiotensin_II_receptor_blocker (6 instances)
 - Anti_arrhythmia_agents
 - Anti_arrhythmia_agent_class_I
 - Anti_arrhythmia_agent_class_II
 - Anti_arrhythmia_agent_class_III (5 instances)



The screenshot displays the Protege software interface for editing the DMOP ontology. The main window title is "DMOP.owl (http://www.e-lico.eu/ontologies/dmo/DMOP/DMOP.owl) - [D:\al\l\00\aa\ELICO\work-elico\ontologies\DMOPver4.4-2010-10-20\DMOP.owl]". The interface includes a menu bar (File, Edit, Ontologies, Reasoner, Tools, Refactor, Tabs, View, Window, Help), a toolbar, and several panels.

Class Hierarchy (Left Panel): Shows a tree structure starting with "Thing", followed by "DM-Entity", "DM-Object", and "Algorithm". Under "Algorithm", there is a large list of subclasses, including "C4.5-FSAAlgorithm", "Chi2TestBasedFSAAlgorithm", "ConsistencyBasedFSAAlgorithm", "CorrelationBasedFSAAlgorithm", "EmbeddedFSAAlgorithm", "EntropyBasedFSAAlgorithm", "FilterFSAAlgorithm", "GeneticFSAAlgorithm", "GAwithFSEvaluatorFSAAlgorithm", "GAwithWrappedLearnerFSAAlgorithm", "HypothesisTestBasedFSAAlgorithm", "InfoGainBasedFSAAlgorithm", "InfoGainRatioBasedFSAAlgorithm", "KolmogorovSmirnovTestBasedFSAAlgorithm", "MultivariateFSAAlgorithm", "NBTree-FSAAlgorithm", "NonParametricTestBasedFSAAlgorithm", "ParametricTestBasedFSAAlgorithm", "ReliefF-FSAAlgorithm", "SVM-RFE-FSAAlgorithm", "SymmetricalUncertaintyBasedFSAAlgorithm", "T-TestBasedFSAAlgorithm", "UnivariateFSAAlgorithm", "WilcoxonTestBasedFSAAlgorithm", "WrapperFSAAlgorithm", "FeatureWeightingAlgorithm", "ModelEvaluationAlgorithm", "ModelProcessingAlgorithm", "ModelPruningAlgorithm", and "ModellingAlgorithm".

Annotations (Top Right Panel): Shows the "Annotations" tab for the selected class "C4.5-FSAAlgorithm". It displays a single annotation: "label 'C4.5-FSAAlgorithm'@".

Description (Bottom Right Panel): Shows the "Description" for "C4.5-FSAAlgorithm". It lists several properties and their values:

- Equivalent classes: (None listed)
- Superclasses:
 - FeatureSelectionAlgorithm
 - hasFeatureEvaluator exactly 1 C4.5-FWAlgorithm
 - hasOptimizationStrategy exactly 1 GreedyBestFirstSearch
 - interactsWithLearnerAs value Embedded
- Inherited anonymous classes:
 - hasOptimizationStrategy exactly 1 DiscreteOptimizationStrategy
 - interactsWithLearnerAs some Filter-Wrapper-Embedded
 - hasFeatureEvaluator exactly 1 FeatureWeightingAlgorithm
 - Algorithm and (addresses some FeatureSelectionTask)
 - specifiesOutputType only IO-Object
 - addresses some CoreDMTask
 - specifiesInputType only IO-Object
 - hasHyperparameter some AlgorithmParameter
- Members: (None listed)
- Keys: (None listed)

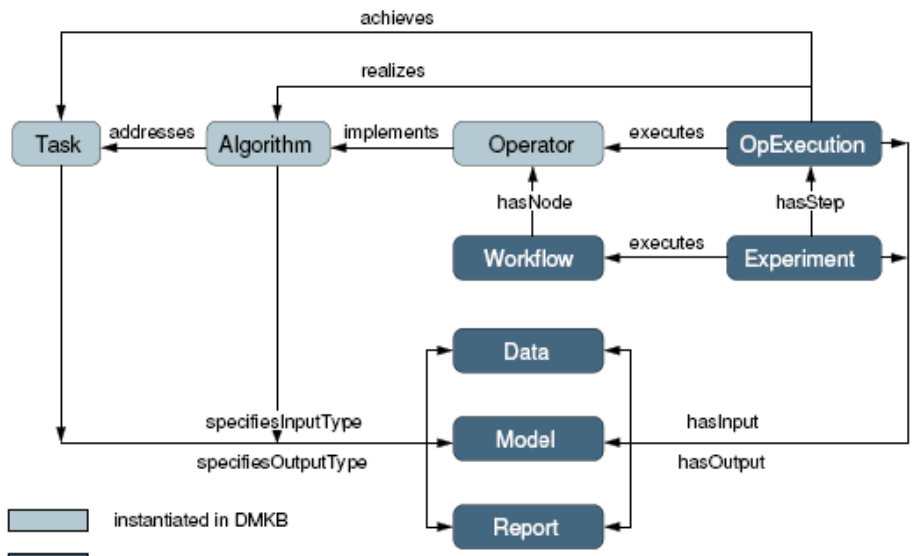
The bottom status bar indicates "No Reasoner set" and "Show Inferences" is checked.

dmop.pdf - Adobe Reader

File Edit View Document Tools Window Help

5 / 22 116% Find

The core DMOP concepts



```

    graph TD
      subgraph DMKB [instantiated in DMKB]
        Task
        Algorithm
        Operator
      end
      subgraph DMEX-DB [instantiated in DMEX-DB]
        OpExecution
        Experiment
        Workflow
        Data
        Model
        Report
      end
      Task -- addresses --> Algorithm
      Algorithm -- implements --> Operator
      Operator -- executes --> OpExecution
      OpExecution -- realizes --> Task
      OpExecution -- hasStep --> Experiment
      Experiment -- executes --> Workflow
      Workflow -- hasNode --> Operator
      Algorithm -- specifiesInputType --> Task
      Algorithm -- specifiesOutputType --> Model
      OpExecution -- hasInput --> Model
      OpExecution -- hasOutput --> Report
  
```

Legend:

- Light blue box: instantiated in DMKB
- Dark blue box: instantiated in DMEX-DB

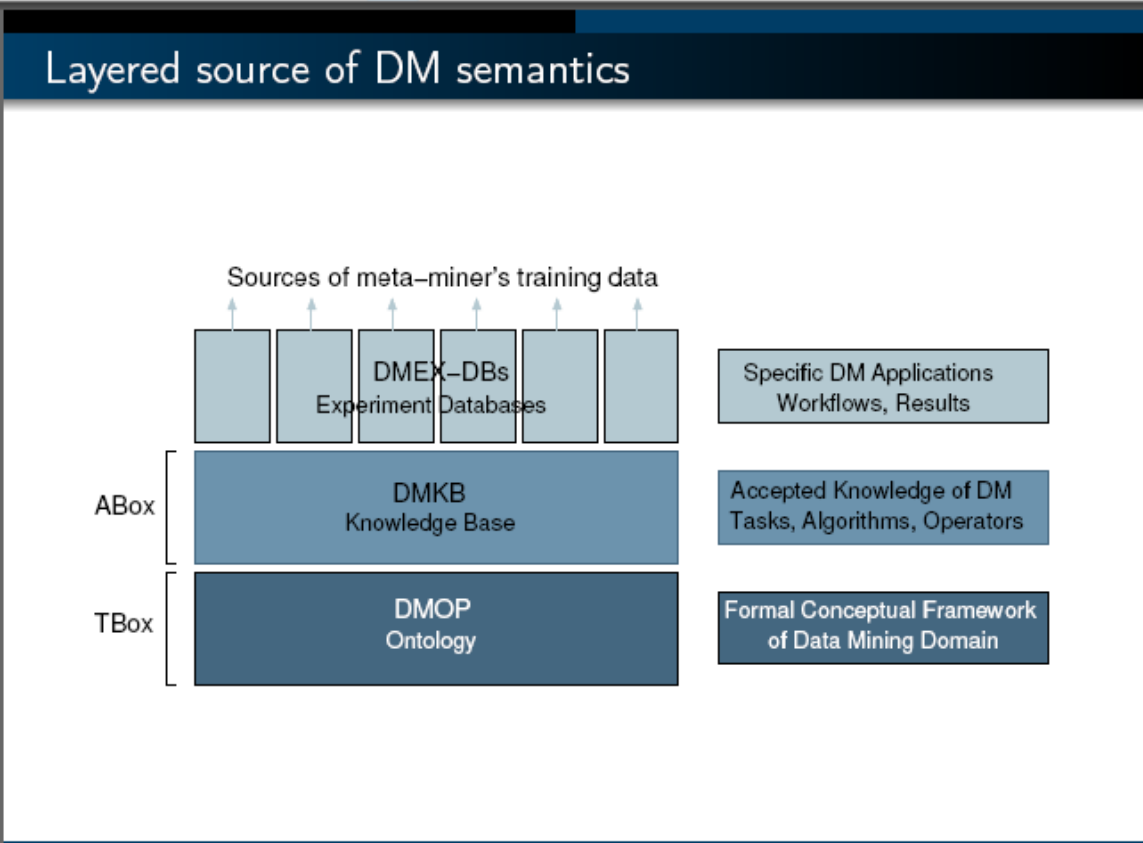
DM Ontology Jamboree-2010 8 Ljubljana, 8 November 2010

dmop.pdf - Adobe Reader

File Edit View Document Tools Window Help

3 / 22 116% Find

Layered source of DM semantics



The diagram illustrates a layered source of DM semantics, structured as follows:

- Sources of meta-miner's training data:** A row of six boxes at the top, with arrows pointing upwards to them. The central boxes are labeled "DMEX-DBs" and "Experiment Databases".
- ABox (Accepted Knowledge):** A large blue box containing "DMKB Knowledge Base". To its right is a text box: "Accepted Knowledge of DM Tasks, Algorithms, Operators".
- TBox (Formal Conceptual Framework):** A large dark blue box containing "DMOP Ontology". To its right is a text box: "Formal Conceptual Framework of Data Mining Domain".
- Specific DM Applications:** A text box on the right side: "Specific DM Applications Workflows, Results".

DM Ontology Jamboree-2010 3 Ljubljana, 8 November 2010



Data Mining ontologija

Trenutno postoji više različitih istraživačkih pristupa:





DMOP - e-LICO projekt (vodi University of Geneva), namjenjena meta učenju, pristup od dolje prema gore.

OntoDM – Institut J.Stefan Ljubljana, konceptualizacija znanja o strojnom učenju, pristup od gore prema dolje.

Expose - University of Leuven, Belgija, namjenjena meta učenju, pristup od gore prema dolje, djelomično integrira i prethodne ontologije.

DMWF - e-LICO projekt (vodi University of Zurich), namjenjena planiranju procesa koji uključuju strojno učenje, sadrži proceduralno znanje pretežno realizirano SWRL jezikom.

Rakorak između očekivanja i stvarnih rezultata

- a) - omogućavaju *jednostavniju implementaciju* i održavanje složenih sustava za obradu podataka i znanja (ekspertni sustavi, analiza podataka). 
- b) - mogu se koristiti u raznim primjenama, osiguravaju njihovu *interoperabilnost*, verifikaciju njihove korektnosti te usporedivost rezultata raznih primjena. 
- c) - korišteni pojmovi moraju biti precizno definirani što je važno za tehničke sustave ali i kao osnova *međuljudske komunikacije* uključujući znanstvene radove, udbenike i stručne procedure (guidelines).  



Zaključci:

- velika očekivanja od prikaza znanja ontologijama
- skromni rezultati njihove primjene
- vrlo zanimljivo područje istraživanja
 - integracija različitih oblika znanja u ontološkom prikazu
 - primjena ontologija u odlučivanju i planiranju
 - razvoj sustava za zaključivanje nad ontologijama
 - ekstrakcija korisnih informacija iz ontologija
 - otkrivanje znanja iz podataka kada je postojeće
 - ljudsko znanje prikazano ontologijama
 - postupci okrivanja znanja pogodni za integraciju u ontologije



Hvala na pažnji !