Knowledge Representation in Big Data
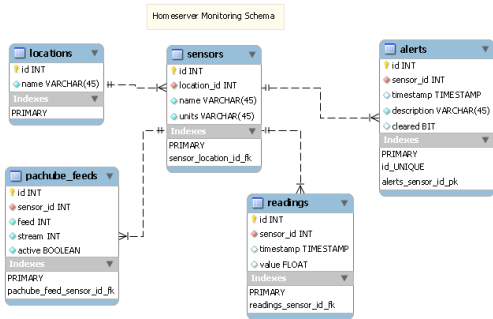
Aleksei Kisllitsin

Definition of Knowledge Representation

- The process of information extraction and knowledge representation, via knowledge processing and analytics to visualization, and practical applications

- It is used to organize the knowledge instances in a graph-based knowledge base

Importance of Knowledge Representation in Big Data

- Ability to represent real-world objects and their relationships in a form that is suitable for processing by software

Homeserver Monitoring Schema

locations
- id INT
- name VARCHAR(45)

Indexes
PRIMARY

sensors
- id INT
- location_id INT
- name VARCHAR(45)
- units VARCHAR(45)

Indexes
PRIMARY
sensor_location_id_fk

alerts
- id INT
- sensor_id INT
- timestamp TIMESTAMP
- description VARCHAR(45)
- cleared BIT

Indexes
PRIMARY
id_UNIQUE
alerts_sensor_id_pk

pachube_feeds
- id INT
- sensor_id INT
- feed INT
- stream INT
- active BOOLEAN

Indexes
PRIMARY
pachube_feed_sensor_id_fk

readings
- id INT
- sensor_id INT
- timestamp TIMESTAMP
- value FLOAT

Indexes
PRIMARY
readings_sensor_id_fk

Knowledge Representation Techniques

1. Semantic web-based knowledge representation
2. Rule-based knowledge representation
3. Logic-based knowledge representation
4. Ontology-based knowledge representation
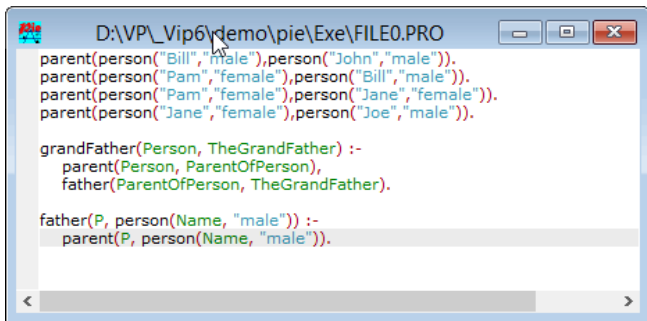
Semantic Networks

- A semantic network is a graph whose nodes represent concepts and whose arcs represent relations between these concepts
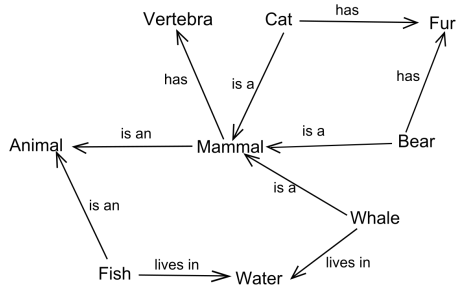
Example of animal's semantic net

Rules

- Rules come in the form of IF-THEN-constructs and allow to express various kinds of complex statements

```prolog
parent(person("Bill","male"),person("John","male")).
parent(person("Pam","female"),person("Bill","male")).
parent(person("Pam","female"),person("Jane","female")).
parent(person("Jane","female"),person("Joe","male")).

grandFather(Person, TheGrandFather) :-
    parent(Person, ParentOfPerson),
    father(ParentOfPerson, TheGrandFather).

father(P, person(Name, "male")) :-
    parent(P, person(Name, "male")).
```

Prolog code snippet

Logic-based

- Allows to describe the domain of interest as consisting of objects, things that have individual identity, and to construct logical formulas around these objects formed by predicates, functions, variables and logical connectives

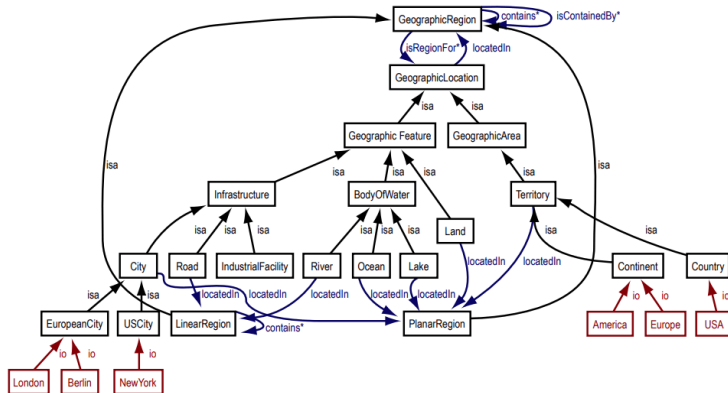| Employee $\xrightarrow{kindOf}$ Person | $\forall x : (Employee(x) \rightarrow Person(x))$ |

| Company $\xrightarrow{books}$ Trip | $\forall x,y : (books(x,y) \rightarrow Company(x) \wedge Trip(y))$
 $\forall x : \exists y : (Trip(x) \rightarrow Company(y) \wedge books(y,x))$ |

Ontology

Challenges

Scalability

Incompleteness

Diversity

Uncertainty

Complexity

Applications

- E-commerce
- Manufacturing
- Telecommunications
- Robotics
- Natural language processing
- Semantic web

Thanks!

Do you have any questions?