

Co to znaczy „wiedzieć” i czy komputery to potrafią?

Krzysztof Goczyła

Teresa Grabowska

Wojciech Waloszek

Michał Zawadzki

Wiedza

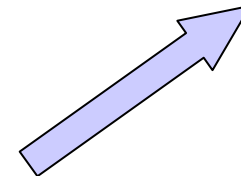
Czym jest *wiedza*?



- Ogół wiadomości zdobytych dzięki uczeniu się.
- Zbiór danych i wszelka inna informacja wydobyta na ich podstawie.



Jak reprezentować wiedzę?



Ontologia jako reprezentacja wiedzy

Ontologia -

formalny opis pojęć występujących w danej dziedzinie („słownik”).

Ontologia

Bezpośrednia specyfikacja konceptualizacji.

Konceptualizacja

Abstrakcyjny model pewnych aspektów świata opisany w terminach konceptów, właściwości konceptów i relacji pomiędzy konceptami.

Bezpośrednia specyfikacja

Model pewnych aspektów świata musi być wyspecyfikowany w jednoznacznym języku nadającym się do przetwarzania zarówno przez komputer, jak i przez człowieka.

Ontologie

Geneza - **ramki** (ang. *frames*) Minsky'ego (1975);
później: model OKBC (Open Knowledge Base Connectivity)

Składniki ontologii wg Minsky'ego:

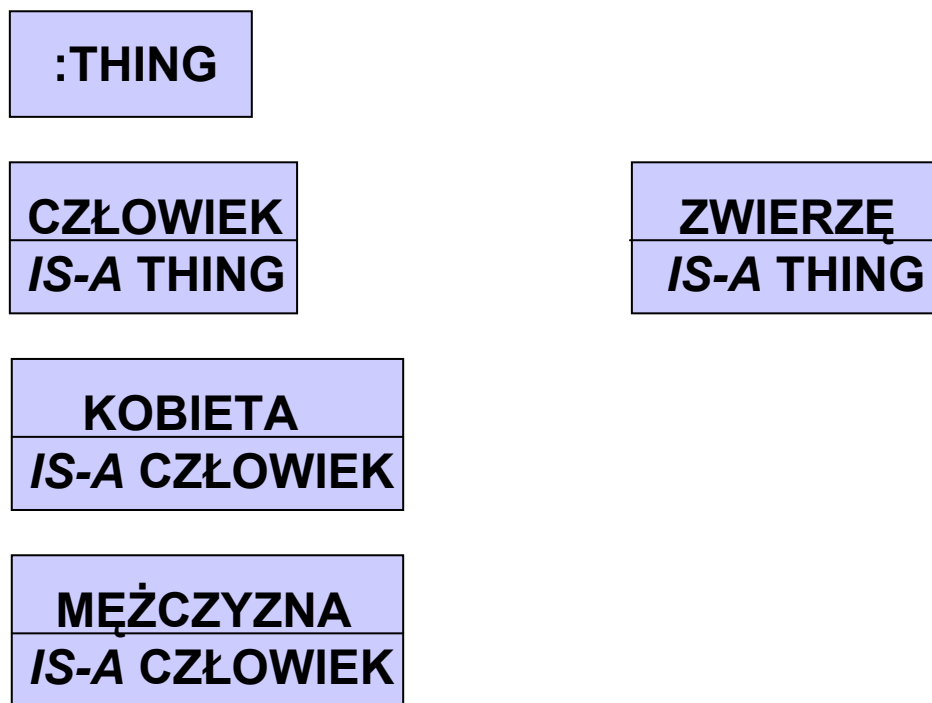
- **klasy** - pojęcia (*koncepty*)
- **klatki** - własności klas opisujące ich cechy i atrybuty (role, właściwości, ang. *slot*)
- **fasety** - własności klatek (np. *ograniczenia* nałożone na klatki)
- **aksjomaty** - dodatkowe ograniczenia wyrażone w języku logiki

Klasy mają swoje **instancje**, które są wystąpieniami klas z określonymi wartościami klatek.

Klasy

Klasy tworzą hierarchię (taksonomię) klas poprzez relację podklasa - nadklasa.

Korzeniem tej hierarchii jest klasa :THING



Klasy i ich instancje

Klasa A jest podklasą klasy B $\Rightarrow$ każda instancja klasy A jest instancją klasy B

THING

CZŁOWIEK
IS-A THING

KOBIETA
IS-A CZŁOWIEK

MĘŻCZYŻNA
IS-A CZŁOWIEK

ZWIERZĘ
IS-A THING

ANNA
INSTANCE_OF KOBIETA



JAN
INSTANCE_OF MĘŻCZYŻNA



MAX
INSTANCE_OF ZWIERZĘ



Klatki i fasety

Klatki są definiowane niezależnie od klas (są pełnoprawnymi ramkami).

Klatki są dołączane do klas i – pośrednio – do obiektów.

Ta sama klatka może być dołączona do wielu klas –
oznacza zawsze to samo.

Fasety - sposób na nałożenie ograniczenie wartości, jakie mogą
przyjmować klatka.

Przykłady faset:

- licznosc (liczba różnych wartości)
- wartość minimalna, wartość maksymalna
- dziedzina (zbiór klas, których instancje mogą być wartościami klatki)
- zakres (zbiór klas, do których można dołączyć klatkę)

Klatki i fasety

CZŁOWIEK
IS-A: THING
PŁEĆ: („Ż”, „M”)

ZWIERZĘ
IS-A: THING
PŁEĆ: („Ż”, „M”)

KOBIETA
IS-A CZŁOWIEK
PŁEĆ: „Ż”

MĘŻCZYZNA
IS-A CZŁOWIEK
PŁEĆ: „M”

ANNA
INSTANCE_OF: KOBIETA



JAN
INSTANCE_OF: MĘŻCZYZNA



MAX
INSTANCE_OF: ZWIERZĘ
PŁEĆ: „M”



Przykład ramek Minsky'ego

NAME	SSAK
SLOT1	SKÓRA: SIERŚĆ
SLOT2	URODZENIE: ŻYWORODNE
SLOT3	PŁEĆ: MĘSKA OR ŻEŃSKA

NAME	PIES
SLOT1	IS-A: SSAK
SLOT2	GŁOS: SZCZEKANIE
SLOT3	NOGI: CARDINALITY: 4

NAME	AZOREK
SLOT1	INSTANCE_OF: PIES
SLOT2	SKÓRA: SIERŚĆ
SLOT3	URODZENIE: ŻYWORODNE
SLOT4	PŁEĆ: MĘSKA
SLOT5	GŁOS: SZCZEKANIE
SLOT6	NOGI: CARDINALITY: 4

NAME	PIES
SLOT1	SKÓRA: SIERŚĆ
SLOT2	URODZENIE: ŻYWORODNE
SLOT3	PŁEĆ: MĘSKA OR ŻEŃSKA
SLOT4	GŁOS: SZCZEKANIE
SLOT5	NOGI: CARDINALITY: 4



INSTANCE_OF

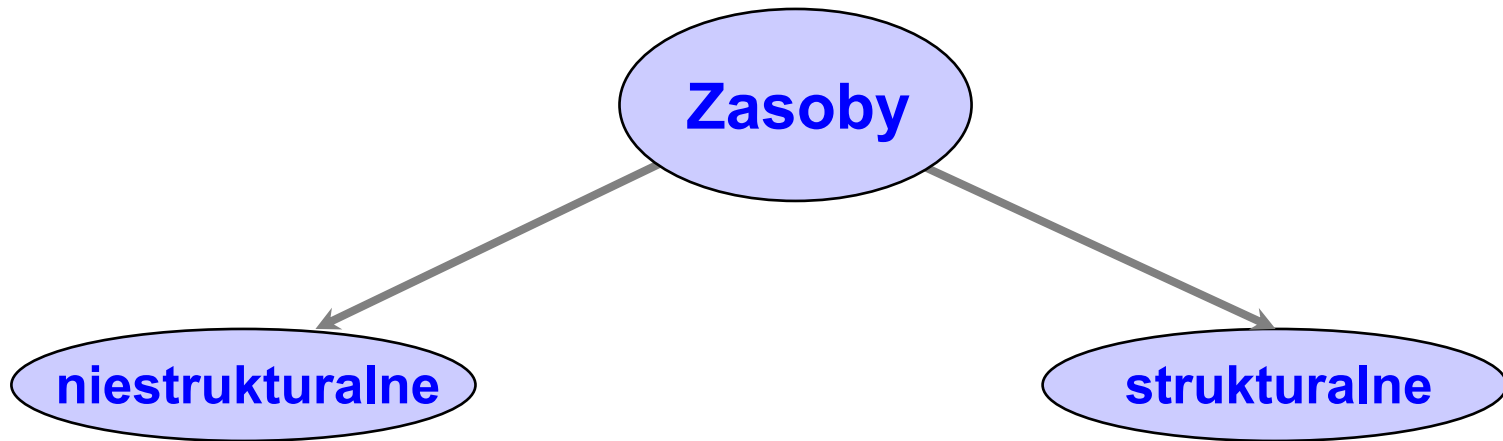
Internet jako składnica danych

- Olbrzymie zasoby informacji
 - *różna struktura*
 - *różna jakość*
- Łatwość dostępu do informacji
(wystarczy mieć przeglądarkę...)
- Łatwość wprowadzania własnych danych
(wystarczy mieć serwer WWW...)

Zalety czy **wady** ?



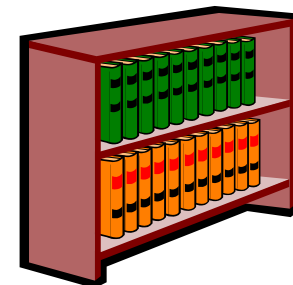
Klasyfikacja zasobów danych



nie istnieje odpowiadający im
schemat danych



istnieje ściśle zdefiniowany
schemat danych



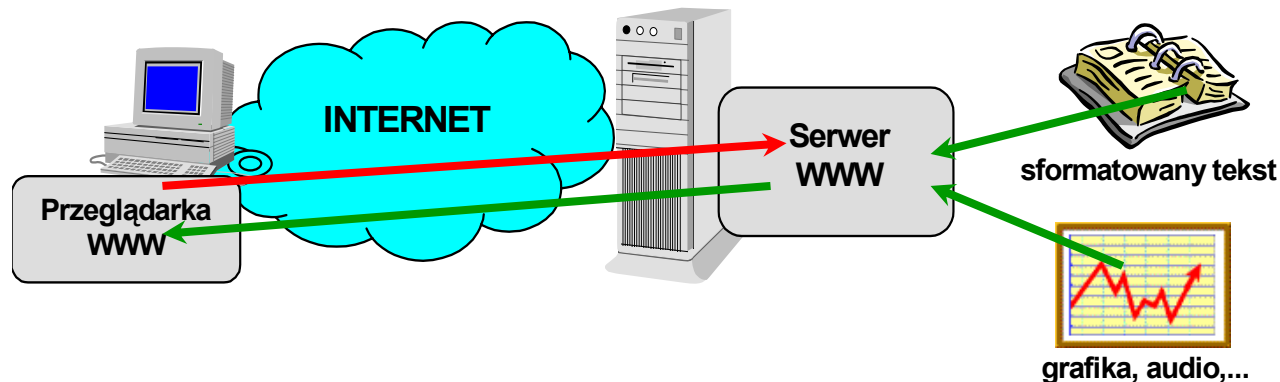
Zasoby niestrukturalne

Statyczne strony (pliki) HTML zawierające

- tekst
- grafikę
- dane audio  , wideo  ...
- odsyłacze do innych stron/plików

Identyfikowane adresem URL (*Uniform Resource Identifier*):

protokół // *komputer.domena* / *plik#miejsce*



Zasoby strukturalne

Dane zgromadzone w bazach danych dostępnych poprzez Internet

Schemat danych + Dane

Dostęp poprzez adres URL rozszerzony o parametry zapytania:

protokół // *komputer.domena* / *procedura* ? *p1=w1&p2=w2* ...



Dynamiczna strona HTML

Zasoby strukturalne

Przykład:

<http://quote.yahoo.com/q?s=intc&d=t>



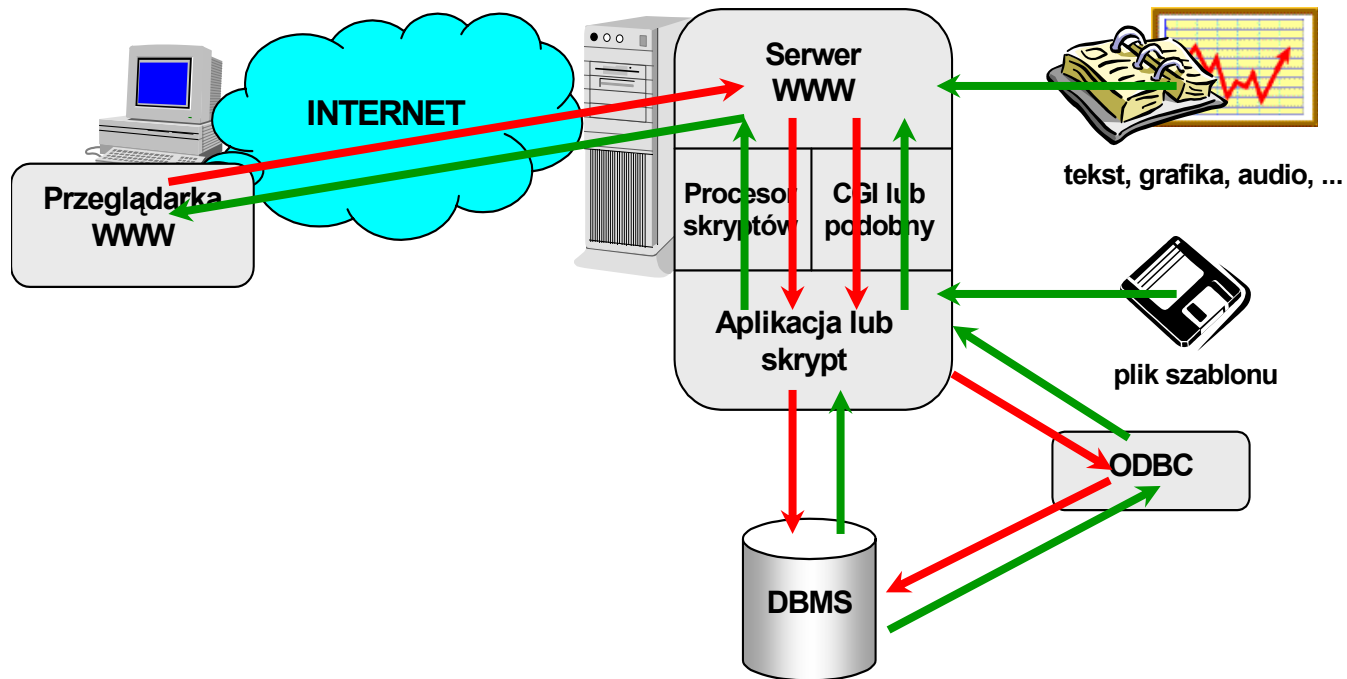
Views: [Basic](#) - [DayWatch](#) - [Performance](#) - [Real-time Mkt](#) - **[Detailed](#)** - [\[Create New View\]](#)

INTEL CORP (NasdaqNM:INTC) - Trade: Choose Brokerage						
Last Trade Jun 5 · 28.18	Change +0.68 (+2.47%)		Prev Cls 27.50	Open 27.60	Volume 41,657,800	
Day's Range 26.89 - 28.20	Bid 28.12	Ask 28.18	P/E 105.77	Mkt Cap 188.4B	Avg Vol 43,997,045	
52-wk Range 18.96 - 36.78	Bid Size 2,000	Ask Size 700	P/S 6.90	Div/Shr 0.08	Div Date Jun 1	
1y Target Est 41.38	EPS (ttm) 0.26	EPS Est 0.70	PEG 2.18	Yield 0.29	Ex-Div May 3	
Chart, Financials, Historical Prices, Insider, Messages, News, Options, Profile, Reports, Research, SEC Filings, Upgrades, more... Compare 22 long-distance plans at a glance. Yahoo! Savings Finder.						

[Add to My Portfolio](#) - [Set Alert](#)

[Download Spreadsheet](#)

Zasoby strukturalne



Zasoby semistrukturalne

Dane semistrukturalne:

- o nieznanym (lub znanym częściowo) schemacie danych
- o strukturze nieregularnej
 - pola opcjonalne
 - alternatywne kombinacje pól
 - pola o nieokreślonej liczności
- niekompletne
- często zawierające błędy i niespójności

Język do opisu i zapisu tego typu danych:

eXtensible Markup Language (XML)

Model wiedzy RDF

Resource Description Framework - model opisu zasobów.

Zasób - wszystko, co można zidentyfikować poprzez Uniform Resource Identifier (URI).

Dokument RDF - opis zasobów i związków pomiędzy nimi.

Schemat RDF - klasy zasobów, typy związków, rodzaje ograniczeń.

Model RDF: zbiór trójek (sieć semantyczna, *Semantic Web*)

	podmiot	predykat	przedmiot
<i>inaczej:</i>	podmiot	orzeczenie	dopełnienie
<i>inaczej:</i>	obiekt	właściwość	wartość

Model wiedzy RDF

Przykład:

Jan | jest_ojcem | Anny
Jan | mówi_że | (Jerzy | jest_ojcem | Anny)

Każdy element trójki jest zasobem identyfikowanym przez URI.

W celu uniknięcia konfliktów nazw wprowadza się pojęcie przestrzeni nazw (*namespace*).

Istnieją przestrzenie nazw standardowych dla RDF: `rdfs`, `rdf`, definiujące podstawowe klasy RDF:

- `rdfs:Resource` - klasa zawierająca wszystkie zasoby
- `rdfs:Class` - klasa zawierająca wszystkie klasy i ich instancje
- `rdf:Property` - klasa zawierająca wszystkie właściwości
- `rdf:Restriction` - klasa zawierająca wszystkie ograniczenia

Inne przestrzenie nazw:

- Dublin Core (dublincore.org)

RDF - przykład



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

OWL (Web Ontology Language)

- Przeznaczony do opisywania i wymiany wiedzy
- Standard rekomendowany przez W3C (WWW Consortium)
- Rozwijany w ramach inicjatywy *Semantic Web*
- Oparty na logice opisowej (ang. Description Logics)



Przykład OWL



```
<owl:Class rdf:ID="Pacjent"/ >
```

```
<owl:Class rdf:ID="PomiarCisnienia"/>
```

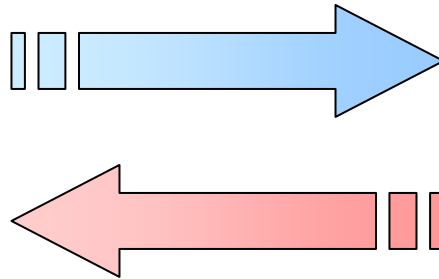
```
<owl:ObjectProperty rdf:ID="maBadanie">
```

```
  <owl:domain rdf:resource="#Pacjent"/>
```

```
  <owl:range rdf:resource="#PomiarCisnienia"/>
```

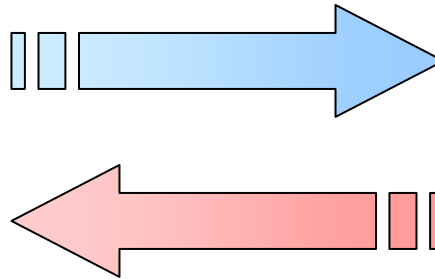
```
</owl:ObjectProperty>
```

Przykład OWL

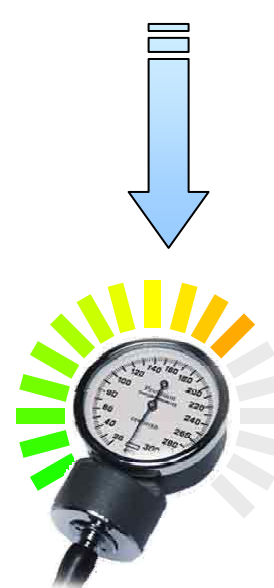


```
<owl:ObjectProperty rdf:ID="dotyczyPacjenta">  
  <owl:domain rdf:resource="#PomiarCisnienia"/>  
  <owl:range rdf:resource="#Pacjent"/>  
  <owl:inverseOf rdf:resource="#maBadanie"/>  
</owl:ObjectProperty>
```

Przykład OWL



Wysoki



Normalny



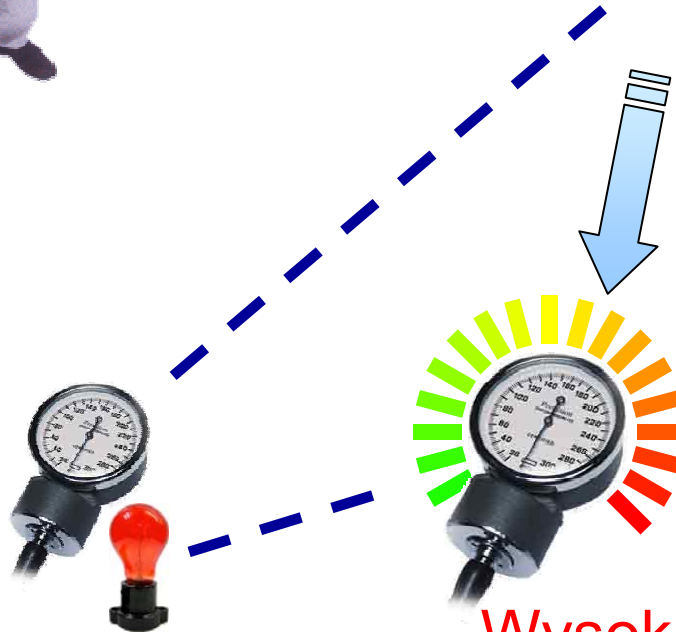
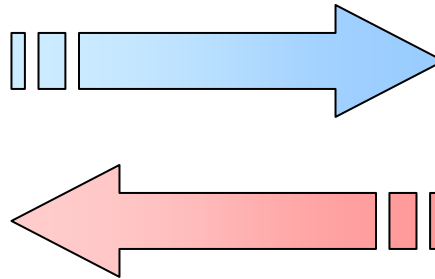
Niski

Przykład OWL

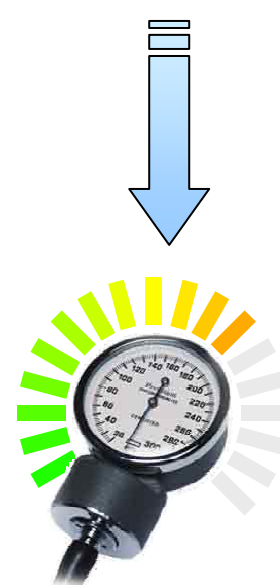
```
<owl:Class rdf:ID="WynikPomiaruCisnienia">  
  <owl:oneOf rdf:parseType="Collection">  
    <owl:Thing rdf:about="#Niski"/>  
    <owl:Thing rdf:about="#Normalny"/>  
    <owl:Thing rdf:about="#Wysoki"/>  
  </owl:oneOf>  
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="maWynik">  
  <owl:domain rdf:resource="#PomiarCisnienia"/>  
  <owl:range rdf:resource="#WynikPomiaruCisnienia"/>  
</owl:ObjectProperty>
```

Przykład OWL



Wysoki



Normalny

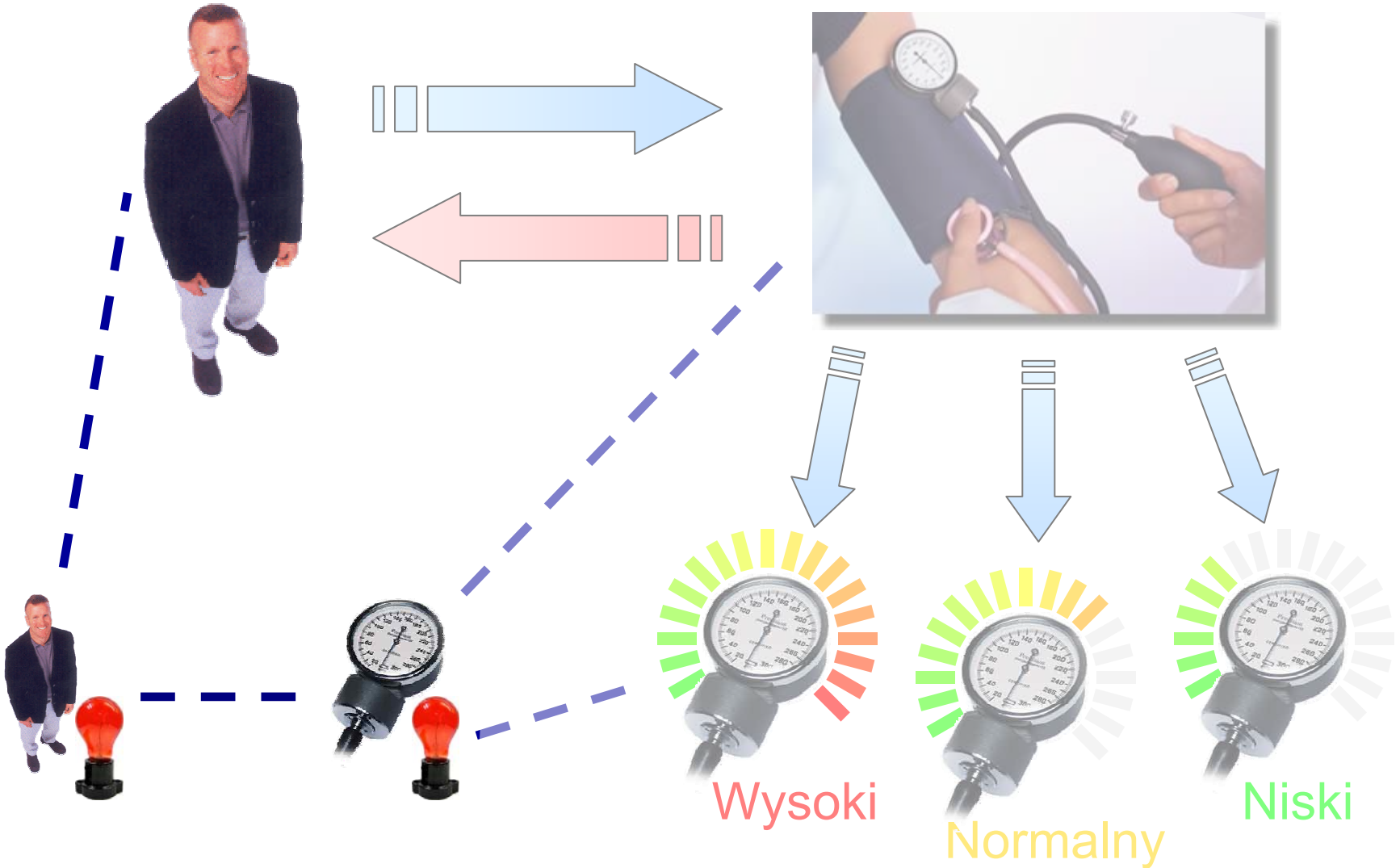


Niski

Przykład OWL

```
<owl:Class rdf:ID="PomiarCisnieniaWykazujacyNadcisnienie">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#WynikPomiaruCisnienia"/>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#maWyniki"/>
      <owl:hasValue rdf:resource="#Wysoki"/>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

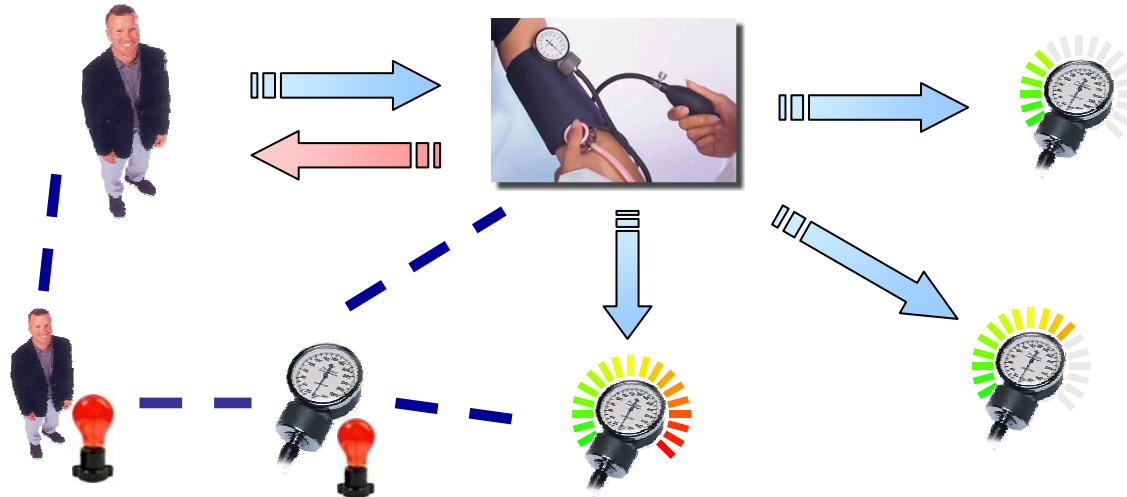
Przykład OWL



Przykład OWL

```
<owl:Class rdf:ID="PacjentZagrozonyNadcisnieniem">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Pacjent"/>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#maBadanie"/>
      <owl:someValuesFrom
        rdf:resource="#PomiarCisnieniaWykazujacyNadcisnienie"/>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

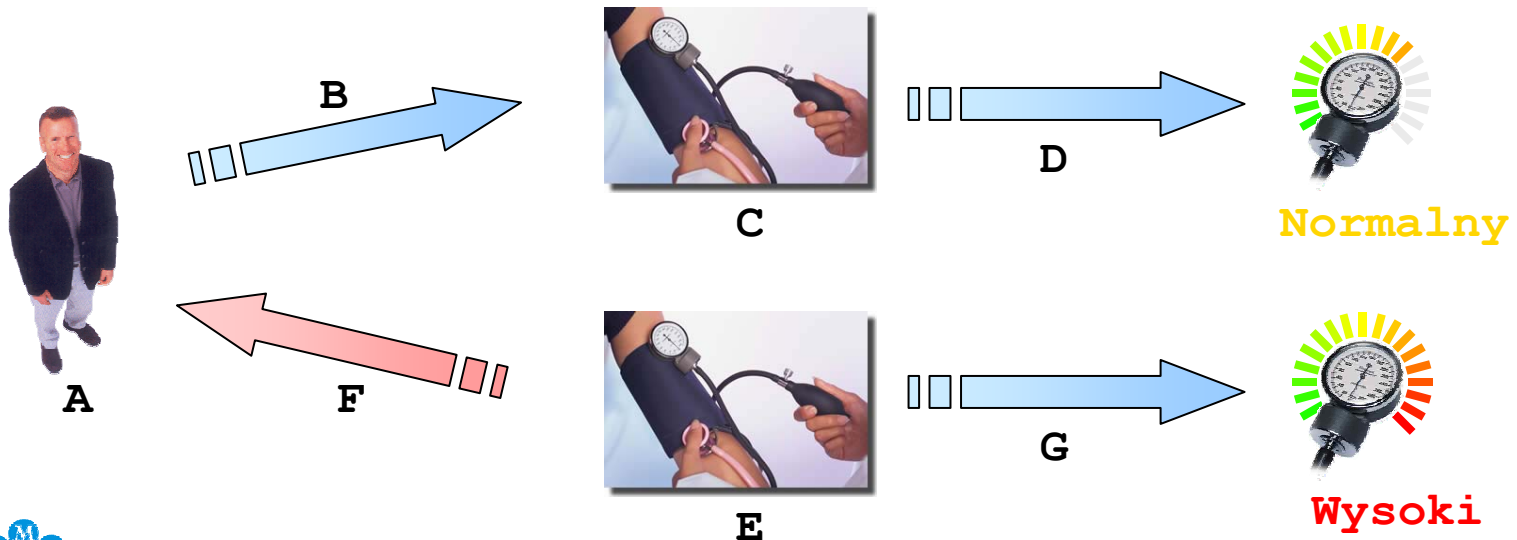
Terminologia



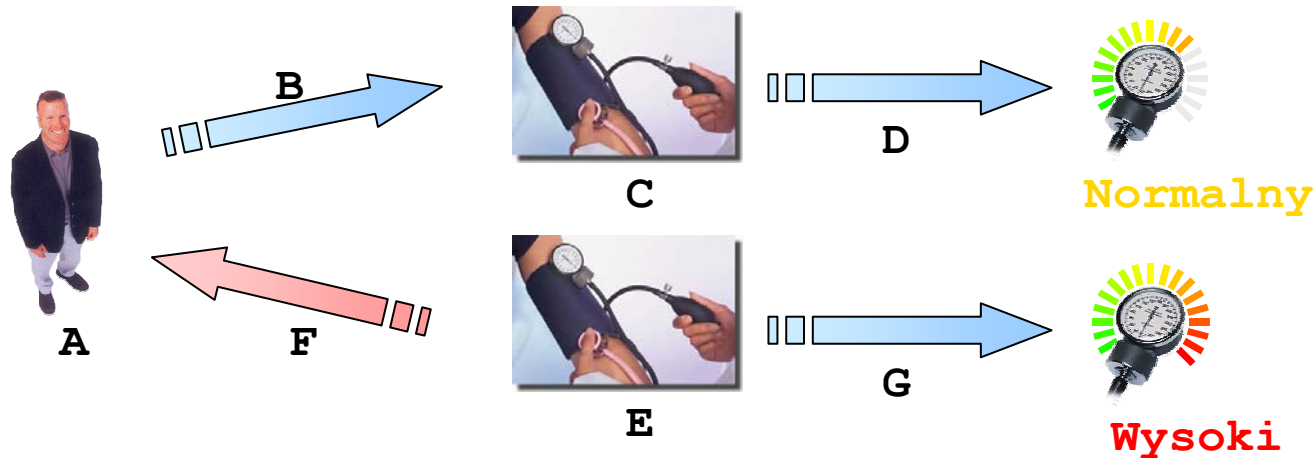
1. $\text{maBadanie}(X, Y) \Leftrightarrow \text{dotyczyPacjenta}(Y, X)$
2. $\text{PomiarCisnieniaWykazujacyNadcisnienie}(X) \Leftrightarrow \text{PomiarCisnienia}(X) \wedge \text{maWynik}(X, \#Wysoki)$
3. $\text{PacjentZagrozonyNadcisnieniem}(X) \Leftrightarrow \text{Pacjent}(X) \wedge \exists Y (\text{maBadanie}(X, Y) \wedge \text{PomiarCisnieniaWykazujacyNadcisnienie}(Y))$

Opis świata

- A. Pacjent(#JanKowalski)
- B. maBadanie(#JanKowalski, #BadanieJK1)
- C. PomiarCisnienia(#BadanieJK1)
- D. maWynik(#BadanieJK1, #Normalny)
- E. PomiarCisnienia(#BadanieJK2)
- F. dotyczyPacjenta(#BadanieJK2, #JanKowalski)
- G. maWynik(#BadanieJK2, #Wysoki)



Opis świata



```

<Pacjent rdf:ID="JanKowalski"/>                                     A
  <maBadanie rdf:resource="#BadanieJK1"/>                           B
</Pacjent>
<PomiarCisnienia rdf:ID="BadanieJK1"/>                               C
  <maWynik rdf:resource="#Normalny"/>                                D
</PomiarCisnienia>
<PomiarCisnienia rdf:ID="BadanieJK2"/>                               E
  <dotyczyPacjenta rdf:resource="#JanKowalski"/>                     F
  <maWynik rdf:resource="#Wysoki"/>                                   G
</PomiarCisnienia>
  
```

Ważne pytanie

Czy Jan Kowalski jest zagrożony nadciśnieniem?

PacjentZagrożonyNadcisnieniem(#JanKowalski)?

Wnioskowanie

I. `maBadanie( #JanKowalski, #BadanieJK2 )`

z F i 1:

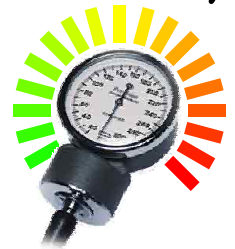
`dotyczyPacjenta( #BadanieJK2, #JanKowalski ) ⇒`
`maBadanie( #JanKowalski, #BadanieJK2 )`



II. `PomiarCisnieniaWykazujacyNadcisnienie( #BadanieJK2 )`

z E, G i 2:

`PomiarCisnienia( #BadanieJK2 ) ∧`
`maWynik( #BadanieJK2, #Wysoki ) ⇒`
`PomiarCisnieniaWykazujacyNadcisnienie( #BadanieJK2 )`



Wnioskowanie

III. PacjentZagrozonyNadcisnieniem(#JanKowalski)

z I, II, A i 3:

Pacjent(#JanKowalski) $\wedge$
mabadanie(#JanKowalski, #BadanieJK2) $\wedge$
PomiarCisnieniaWykazujacyNadcisnienie(#BadanieJK2) $\Rightarrow$
PacjentZagrozonyNadcisnieniem(#JanKowalski)

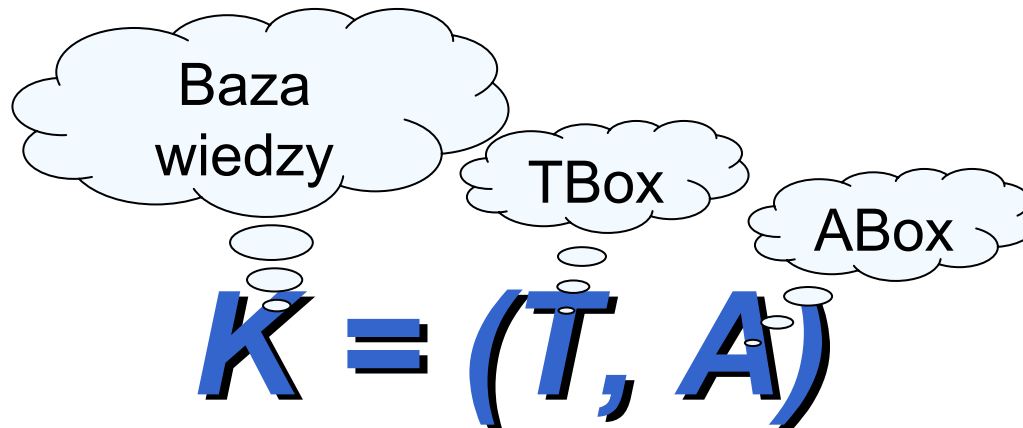


Wnioskowanie



Jan Kowalski jest zagrożony nadciśnieniem!

Baza wiedzy - składniki



TBox (wiedza terminologiczna) – zbiór conceptów i relacji pomiędzy nimi (zwanymi *rolami*).

ABox (wiedza asercjonalna) – zbiór wystąpień (*instancji*) conceptów i ról.

Baza wiedzy

TBox

hasUncle
hasSister
hasHusband
hasAunt
hasChild
hasSon $\sqsubseteq$ hasChild

Role

Man $\sqsubseteq$ Person
Woman $\sqsubseteq$ Person
Child $\sqsubseteq$ Person

Koncepty

ABox

Man(JohnSmith)
Man(HarryBlack)
Woman(MaryBlack)
Child(BillSmith)
hasSister(JohnSmith, MaryBlack)



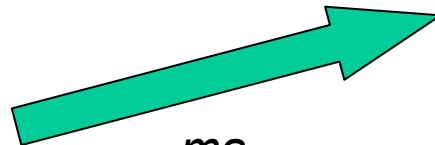
Świat ludzi i zwierząt

- Świat dzieli się na *Ludzi* i *Zwierzęta*
- *Zwierzęta* to *Psy* i *Koty*
- *Człowiek* może mieć *Zwierzę* (wówczas staje się *Właścicielem Psa* lub *Właścicielem Kota*)
- Między *Kotami* i *Psami* trwają odwieczne walki:
 - *Kot* nie lubi *Psa* ani *Właściciela Psa*
 - *Pies* nie lubi *Kota* ani *Właściciela Kota*

Ala ma ...



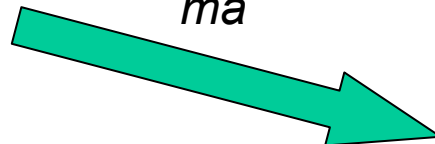
Ala



ma



Max



ma

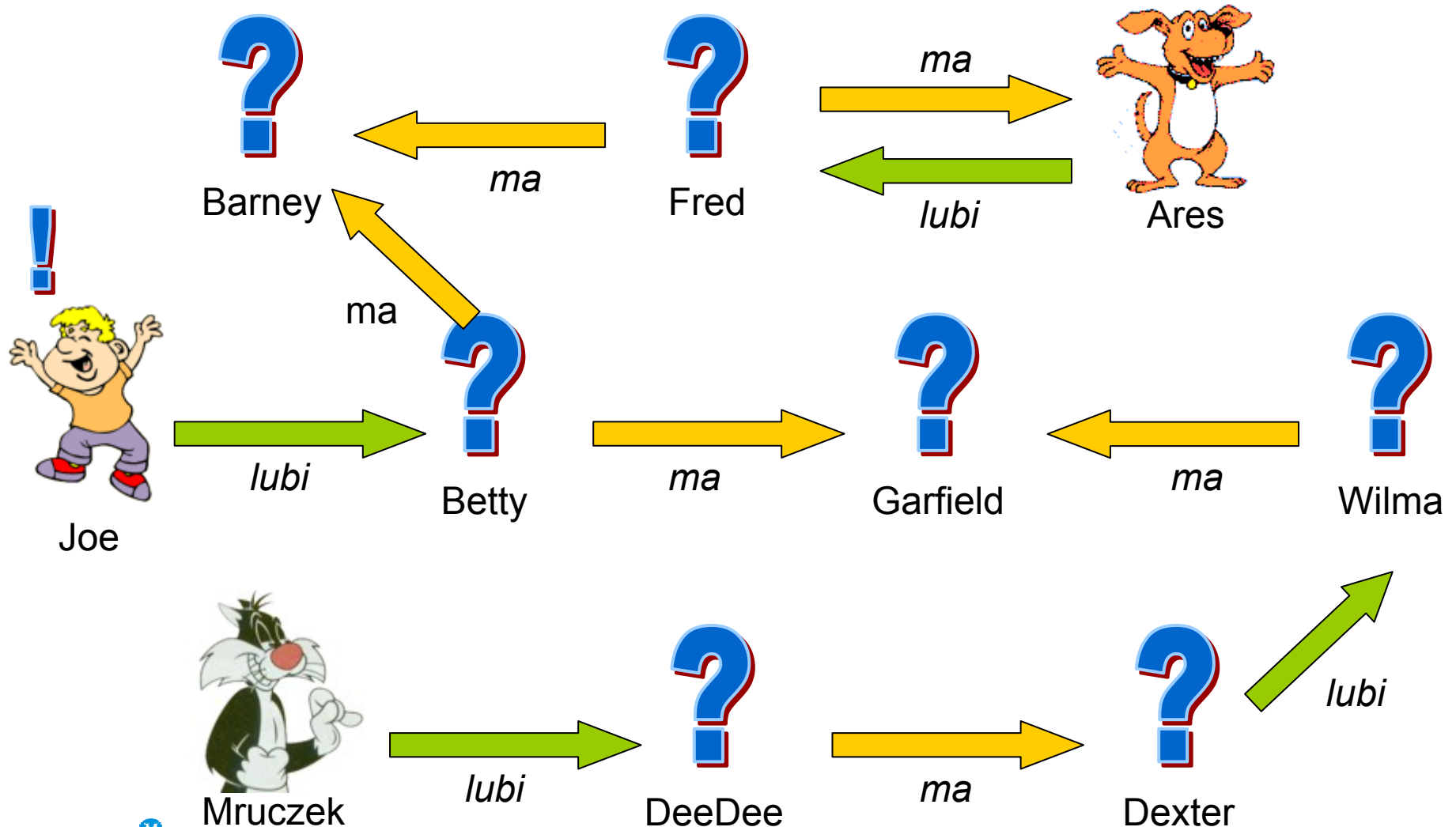


As

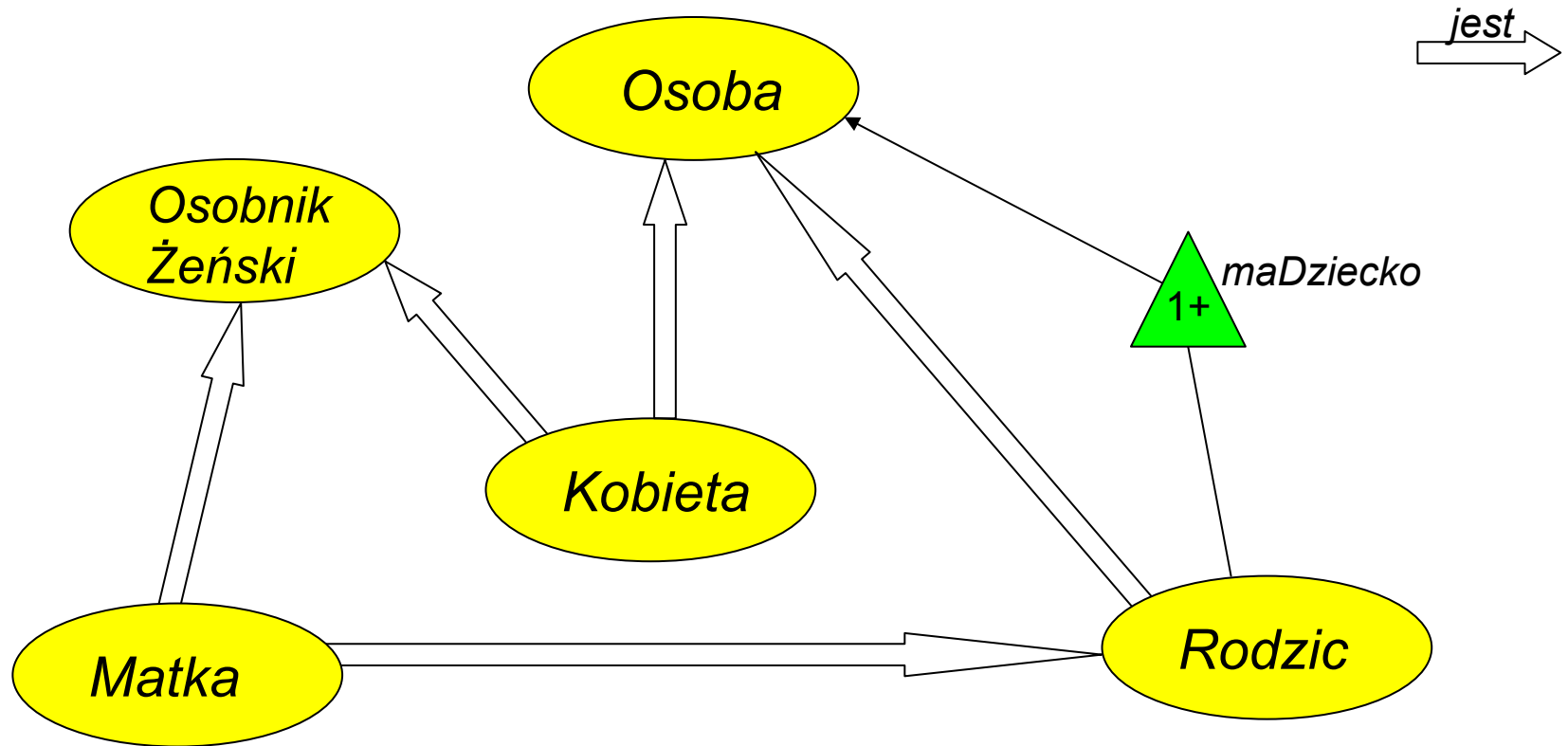


lubi

Zwierzyniec

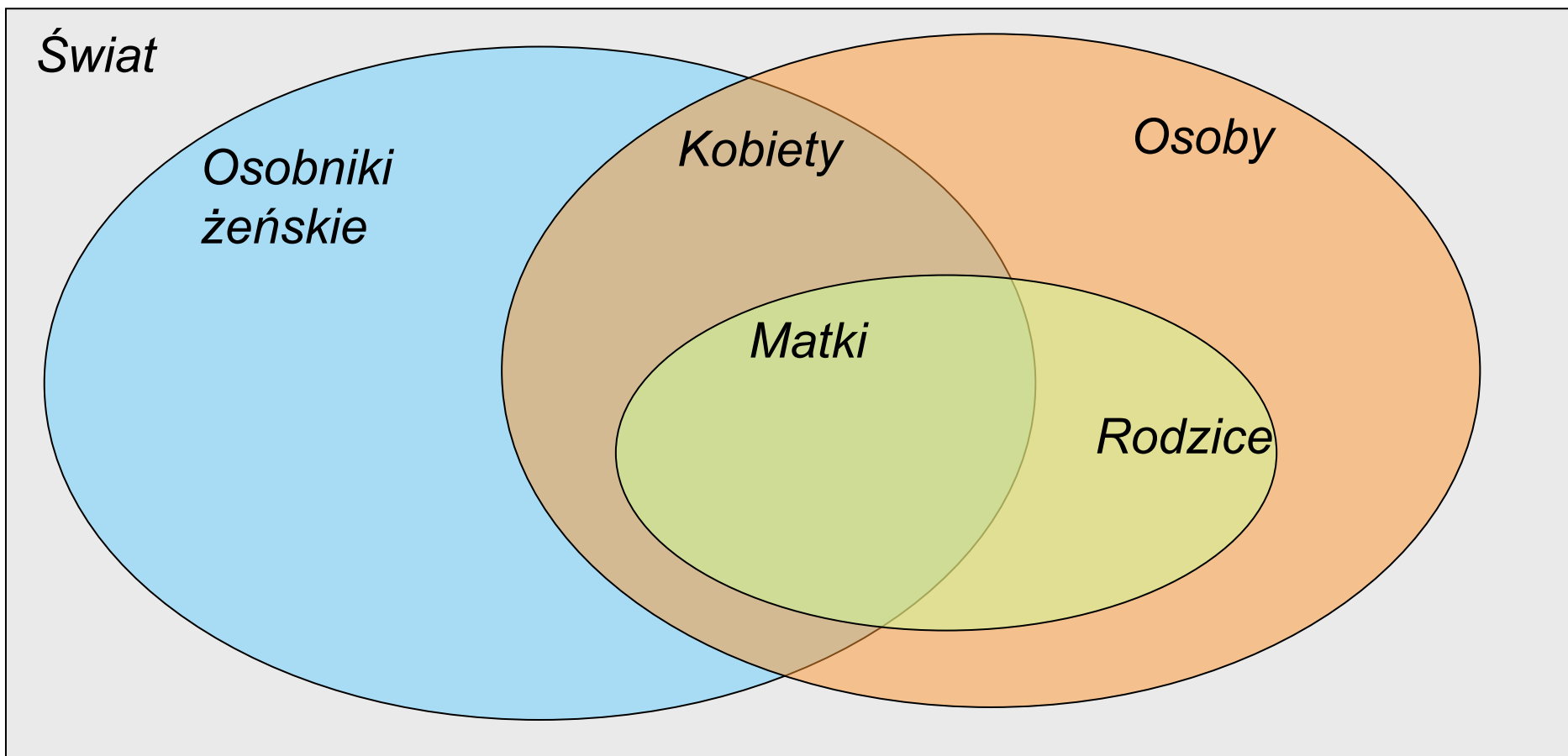


Przykładowa sieć semantyczna

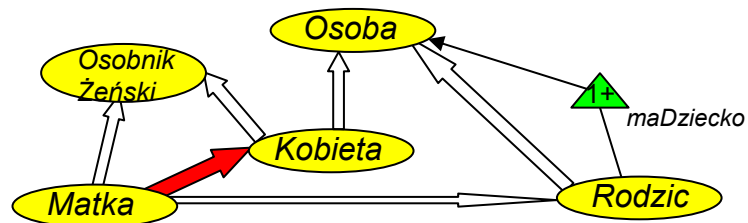


Jaki jest związek pomiędzy *Matką* a *Kobietą*?

Przykład wnioskowania



Mapa wiedzy



Przykład bazy wiedzy

Terminologia (TBox):

Kobieta

Mężczyzna

Kobieta $\equiv \neg$ *Mężczyzna*

maPrzyjaciela

kocha

Świat (ABox):

Kobieta (Susan)

Mężczyzna (Peter)

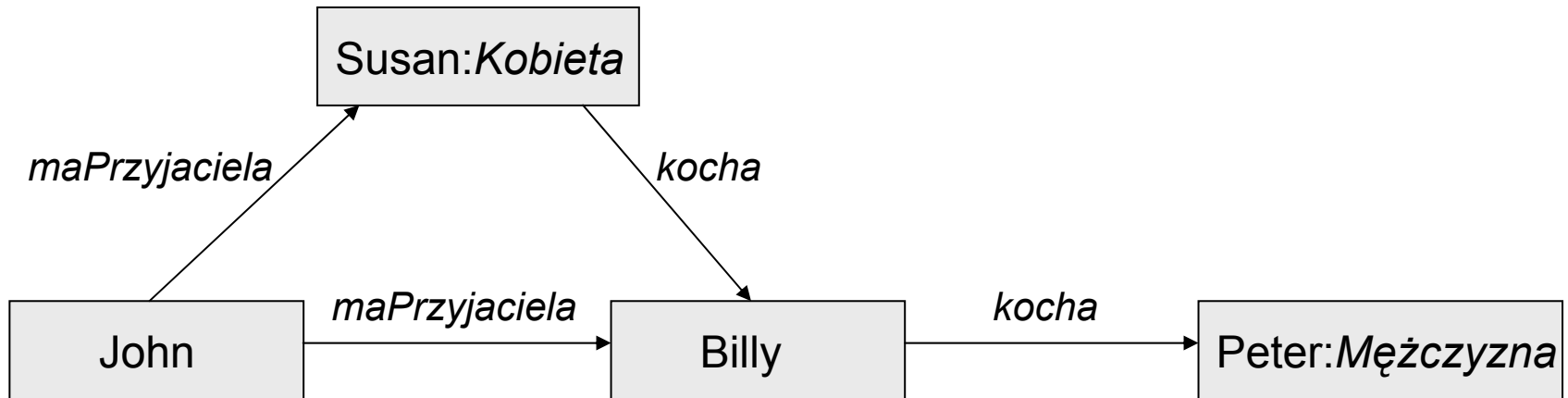
maPrzyjaciela (John, Susan)

maPrzyjaciela (John, Billy)

kocha (Susan, Billy)

kocha (Billy, Peter)

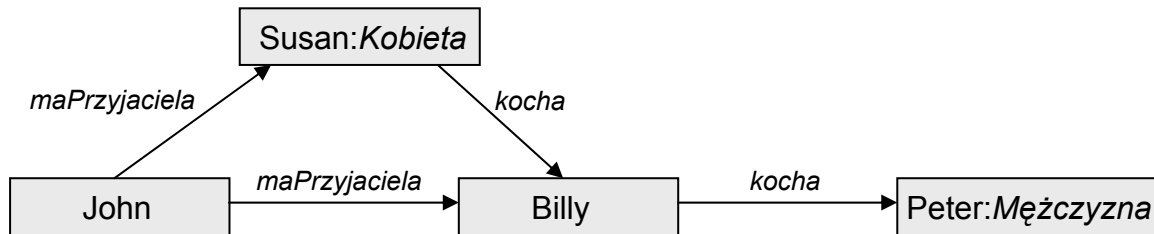
Przykład bazy wiedzy



Problem:

Czy John ma przyjaciółkę, która kocha mężczyznę?

Świat zamknięty (baza danych)



Kobiety

imię
Susan

Mężczyźni

imię
Peter

Przyjaciele

kto	kogo
John	Susan
John	Billy

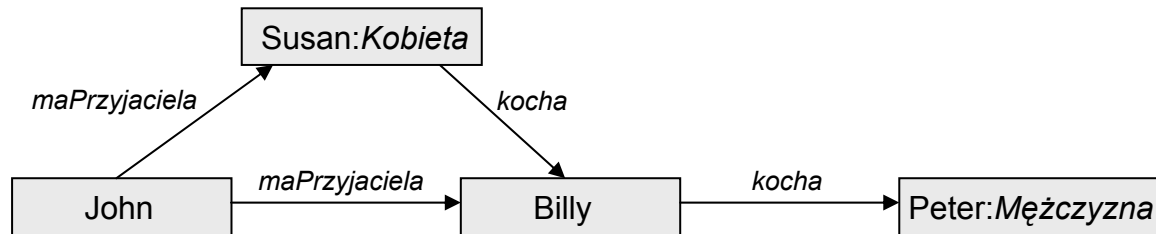
Kochankowie

kto	kogo
Susan	Billy
Billy	Peter

Czy John ma przyjaciółkę, która kocha mężczyznę?

NIE!

Świat otwarty (baza wiedzy)



Czy John ma przyjaciółkę, która kocha mężczyznę?

Billy jest albo mężczyzną, albo kobietą.

Jeśli Billy jest mężczyzną:
(tą przyjaciółką jest Susan)

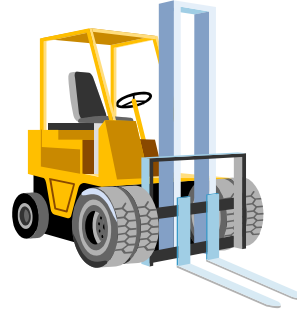
TAK!

Jeśli Billy jest kobietą:
(tą przyjaciółką jest Billy)

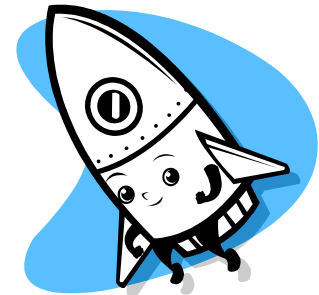
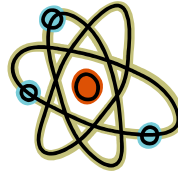
TAK!

Zastosowania

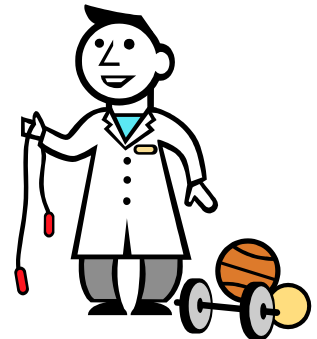
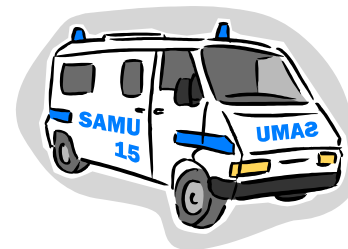
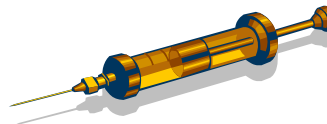
- W przemyśle



- W nauce



- W medycynie



Materiały źródłowe

XML, XMLSchema

www.w3.org



RDF, RDFSchema

www.w3.org



OWL Web Ontology Language Guide

www.w3.org



Protégé

protege.stanford.edu

(dokumentacja+program+dodatki+ontologie)

Semantic Web

www.w3.org/2001/sw

Dziękuję za uwagę.

