

Semantic Web w relacyjnych bazach danych – realizacja zapytań

Jarosław Bąk

JUG Poznań 09.11.2010

**Instytut Automatyki i Inżynierii Informatycznej
Politechnika Poznańska
Jaroslaw.Bak@put.poznan.pl**

Agenda

1. Co to jest Semantyczna Sieć WWW?
2. Najważniejsze technologie
3. Narzędzia dla Semantycznej Sieci
4. Semantic Web w bazach danych
5. Przykład

Co to jest Semantic Web?

Co to jest Semantic Web?

- **Wizja** Sieci WWW wg Tima Bernersa-Lee (1998)
- Przetwarzanie danych w sposób adekwatny do ich **znaczenia** (semantyki)
- Dane są dokładnie opisane (**meta dane**)
- Ekstrakcja informacji z różnych źródeł i ich **automatyczna integracja**
- „**Rozumienie**” danych przez maszyny (**inteligentne przetwarzanie**) - agenci

Co to jest Semantic Web?

- Zarządzanie **wiarygodną** informacją a nie samymi danymi
- **Wywodzenie** nowych informacji poprzez wnioskowanie
- **Formalizacja** wiedzy dziedzinowej – ontologia
- **Zadawanie zapytań** w języku naturalnym

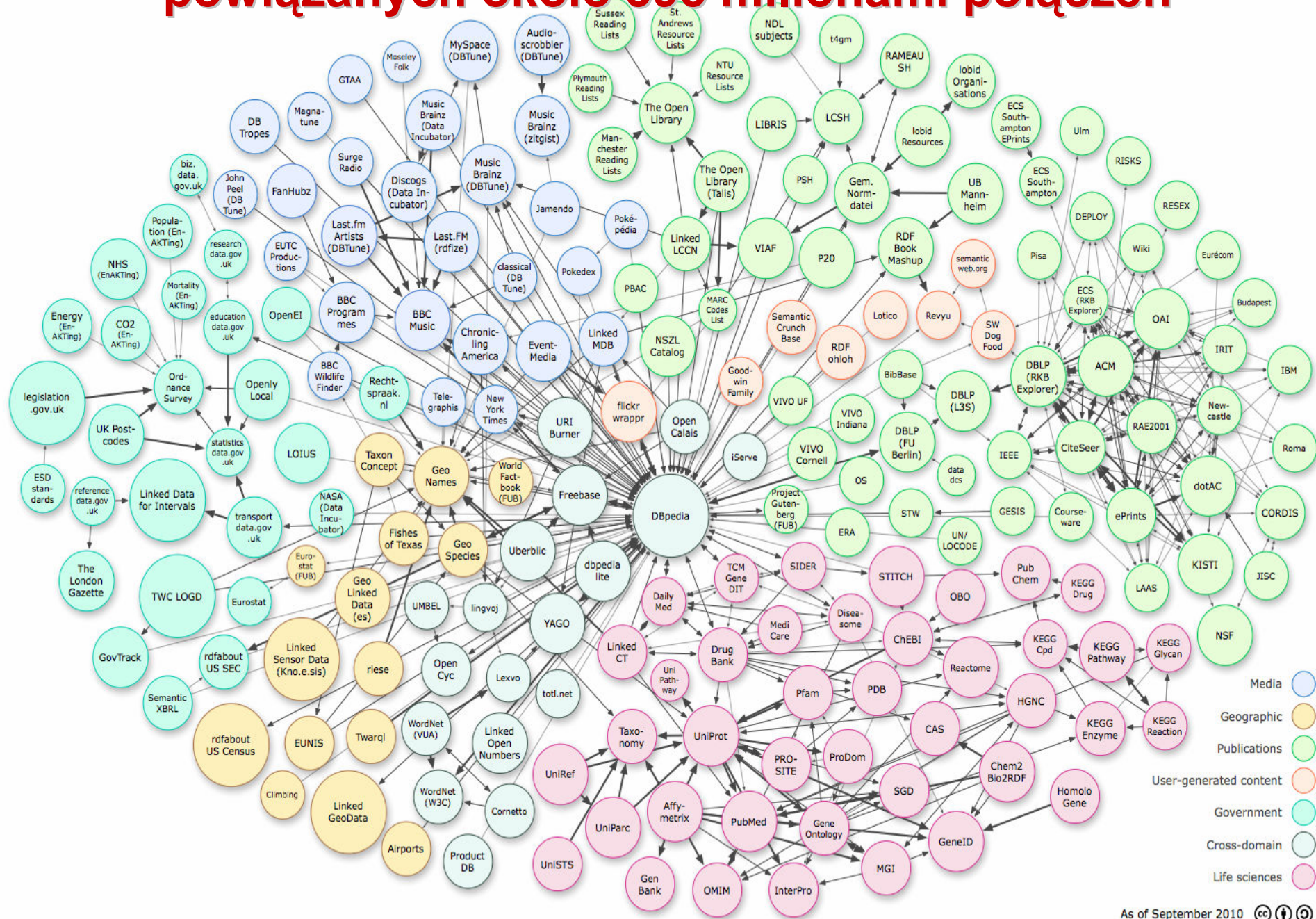
Semantic Web dzisiaj

- **Rozproszona** sieć danych, w której każda informacja może być powiązana w sposób logiczny z dowolną inną informacją
- Największa **formalna** baza danych istniejąca na świecie a jednocześnie najbardziej „zabałaganiona”
- Zbiór **standardów** W3C: XML, XML Schema, RDF, RDFS, OWL, OWL2, F-Logic, SPARQL, SWRL, RIF, RuleML, SBVR i inne
- Motywacja do rozwoju aplikacji z zakresu sztucznej inteligencji – **skalowalność**
- Najważniejszy element **Web 3.0** (Lee, 2006)

Zastosowania

- Semantyczne **wyszukiwarki**: PowerSet, Yahoo! SearchMonkey, Google Rich Snippets i inne
- Semantyczne **Wiki**: Semantic MediaWiki, BiomedGT, VisWiki
- **Bazy wiedzy**: Freebase, **Linking Open Data**, DBpedia
- Rzetelność i **wiarygodność** uzyskanych informacji
- Użytkownik otrzymuje dokładnie to czego oczekuje !!!
- Inne.. z których obecnie sobie w ogóle nie zdajemy sprawy (Facebook, NK, IM...)

LOD - 25 miliardów trójek RDF, powiązanych około 395 milionami połączeń



Najważniejsze technologie Semantycznej Sieci WWW

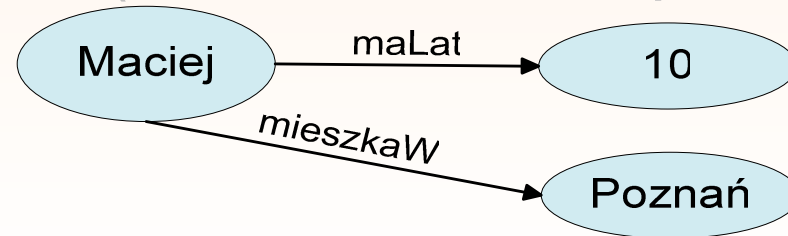
Najważniejsze technologie - ontologia

- Ontologia zajmuje się **opisywaniem** danych (meta dane)
- Charakteryzuje **fragment rzeczywistości**: mniej lub bardziej szczegółowy (ontologie fundacjonistyczne, dziedzinowe)
- Pozwala na **kategoryzację i hierarchizację** pojęć (taksonomia): zwierzę <- ssak <- człowiek
- Taksonomia pojęć wprowadza **relację dziedziczenia** (podobnie jak w językach programowania obiektowego)
- Ontologia pozwala **wyrazić model relacyjny** i go **wzbogacić** na poziomie **semantycznym**
- Dołączenie pojęć ontologicznych do danych relacyjnych (lub dowolnych innych) pozwala na uzyskanie **danych z jawnie wyspecyfikowaną semantyką**
- Reguły **zwiększają siłę wyrazu** ontologii oraz możliwości analizy właściwości semantycznych przetwarzanych danych

Najważniejsze technologie

- Reprezentacja danych – RDF (Resource Description Framework) – **trójki**:

Maciej maLat 10 .
Maciej mieszkaW Poznań .



- Język RDFS (Resource Description Framework Schema):

- wprowadza pojęcie klasy oraz relacji,
- pozwala na definiowanie ograniczeń dla klas i relacji
- umożliwia budowanie **taksonomii** (hierarchii) klas i relacji

```
Maciej rdf:type Człowiek .  
Człowiek rdf:type rdfs:Class .  
Człowiek rdfs:subClassOf Zwierzę .
```

- **Formalizacja wiedzy** – OWL (Web Ontology Language):
zbiór klas, relacji oraz danych (instancji klas i relacji) o
większej sile wyrazu niż RDFS

Najważniejsze technologie

- Zapis **reguł wnioskujących** – SWRL (Semantic Web Rule Language) – łączenie klas i relacji w (prawie) dowolny sposób

```
Człowiek(?x), maLat(?x, ?y), ?y ≥ 18  
-> DorosłyCzłowiek(?x)
```

- Język **zapytań** bazujący na trójkach RDF – SPARQL (SPARQL Protocol and RDF Query Language)

```
SELECT ?x, ?y WHERE  
{  
  ?x maLat ?y .  
  ?x mieszkaW Poznań  
}
```

```
Maciej maLat 10 .  
Maciej mieszkaW Poznań .
```

```
?x = Maciej, ?y = 10
```

Narzędzia dla Semantycznej Sieci

Narzędzia wnioskujące

- **Pellet** (dwie licencje)
- KAON2 (wersja komercyjna – OntoBroker z Ontoprise)
- **Jess** (Java Expert System Shell) – możliwa licencja akademicka
- Jena2 – darmowy framework HP do tworzenia elementów Semantic Web
- RacerPro – możliwa licencja akademicka
- **Oracle** 11g RDFS/OWL – system wnioskowania oparty na bazie Oracle
- AllegroGraph – komercyjna **baza danych RDF** umożliwiająca wnioskowanie (głównie RDFS i parę aksjomatów OWL) oraz zadawanie zapytań
- **D2RQ** – platforma do zadawania zapytań do relacyjnych baz danych w języku SPARQL. Dane traktowane jako wirtualne grafy RDF
- Wiele innych

Pellet

- Narzędzie posiada **dwie** licencje: komercyjną i Open Source (AGPL v. 3)
- Wsparcie **wnioskowania** dla języków: RDFS, OWL, OWL2, SWRL
- Dedykowane **interfejsy** dostępu różnego rodzaju: DIG, API, Jena i inne
- Sprawdzanie poprawności i spójności ontologii
- Wiele, wiele innych – kwestie techniczne i formalne

Jess

- Java Expert System Shell
- Licencja komercyjna (możliwa również **akademicka**)
- Wersja testowa na 30 dni
- System **eksperski** zaimplementowany w Javie
- Wnioskowanie **w przód** oraz w tył za pomocą reguł: **IF** (...) **=>** **THEN** (...)
- Szereg funkcji oraz metod do zarządzania silnikiem z poziomu Javy
- Mała **wielkość** < 1MB
- Możliwość **pełnej integracji** z aplikacjami w Javie
- Zastosowanie w **Navy's DDG 1000** destroyer (jedno z nielicznych oficjalnych zastosowań)



OWL-API

- Narzędzie do obsługi plików **OWL/OWL2**
- Różnego rodzaju **parseery**: RDF/XML, OWL/XML, Turtle i inne,
- Implementacja w Javie
- Wykorzystywane w wielu **projektach**
- Poza narzędziem Jena2 praktycznie **najlepsza** biblioteka do radzenia sobie z plikami zawierającymi ontologię
- Wsparcie dla wielu narzędzi wnioskujących: Pellet, FaCT++, HermiT, Racer
- Interfejs dedykowany dla **Protégé**

Protégé

- **Edytor** ontologii
- Open Source
- Rozwijane przez Stanford
- Dostępne na **każdą** platformę
- Implementacja w Javie
- Zbudowany na **Equinox** (OSGI)
- Najnowsza wersja (4.x.x) posiada wsparcie dla OWL/OWL2, SWRL, DIG, Pellet, FaCT++

Semantic Web w bazach danych

Semantyka baz danych

- Semantyka zawarta w schemacie bazy danych znana **tylko** projektantowi bazy (ewentualnie czytelnikowi dokumentacji 😊)
- Semantyka bazy „**rozbita**” w celach wydajnościowych (postać normalna itd.)
- Problem **identyfikatorów** (unikalne w całej bazie)

- Bazy **RDF**
- Różne metody **osadzania** trójek RDF w bazach danych (nie tylko relacyjnych)

Metody dostępu

- Bazy danych **RDF** – zoptymalizowane pod kątem zapisywania grafów do struktur relacyjnych (Jena2, AllegroGraph i inne)
- Statyczne:
 - **DataMaster** (mapowanie na stałe – opłacalne tylko dla małych i nie zmieniających się baz danych)
 - **Oracle** (ontologia w bazie - każda nowa dana wymaga uruchomienia jeszcze raz procesu wnioskowania)
- Dynamiczne:
 - **KAON2** (mało wydajne, żmudny proces tworzenia odwzorowań)
 - **D2RQ** (wydajny, zapytania w SPARQL, wsparcie dla OWL i RDF(S))
 - **SDL** (wydajny, zapytania w postaci grafów RDF, wsparcie dla OWL/OWL2, RDF(S), SWRL, brak SPARQL)

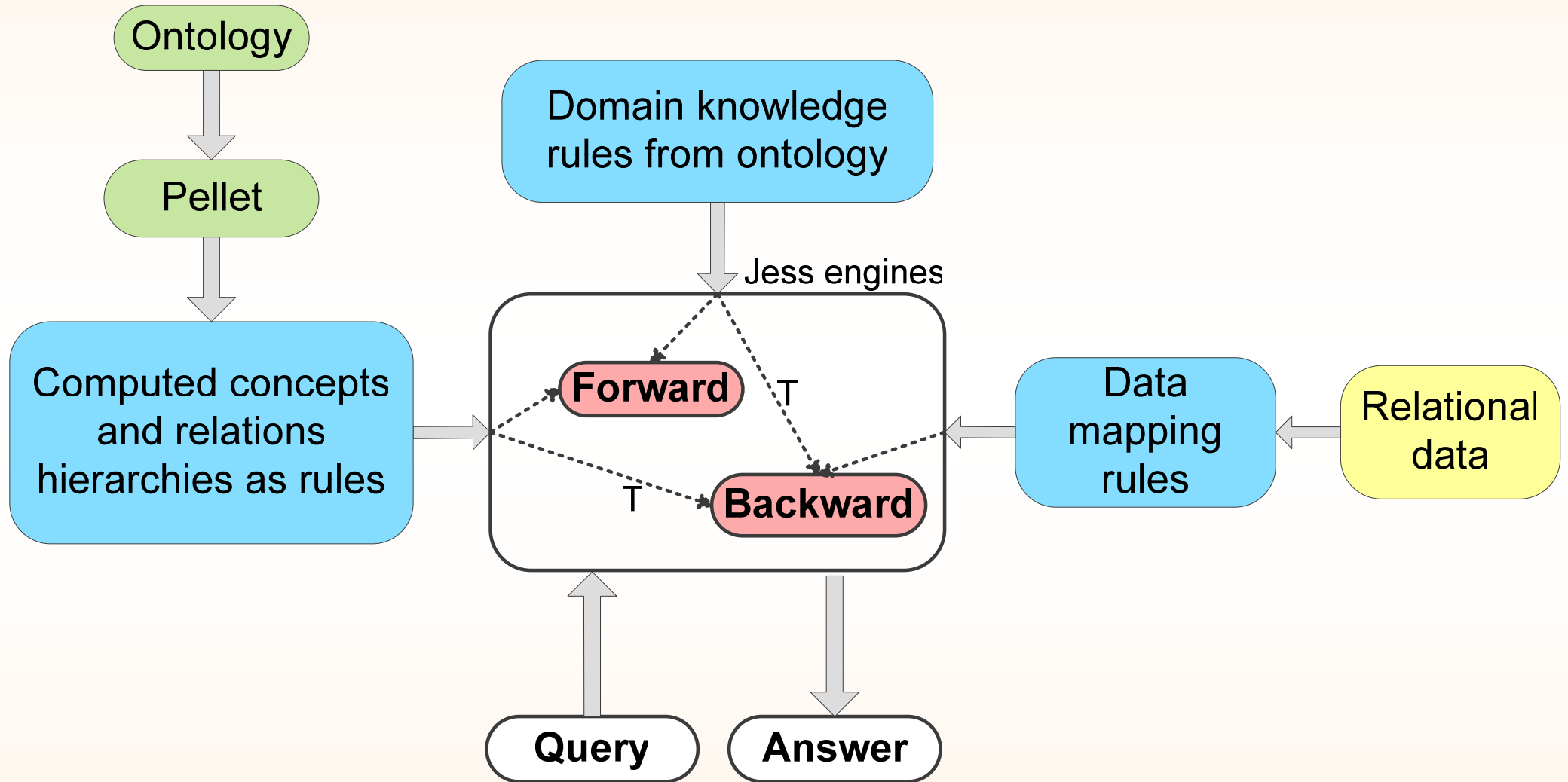
SDL

- Semantic Data Library
- Brak oficjalnej licencji, ale **dostępne**
- Implementacja w Javie
- Moduły **API** oraz **GUI**
- **Integracja** ontologii (OWL/OWL2 + SWRL), systemu regułowego (Jess) oraz baz danych (MS SQL)
- Ciągły **rozwój**
- Modyfikacja **procesów wnioskowania** w przód, wstecz oraz algorytmy hybrydowe
- Metody **automatycznej** integracji
- Szereg funkcji dostępnych dla narzędzia **Jess**
- Planowane wsparcie dla JBoss **Drools**

SDL - funkcjonalności

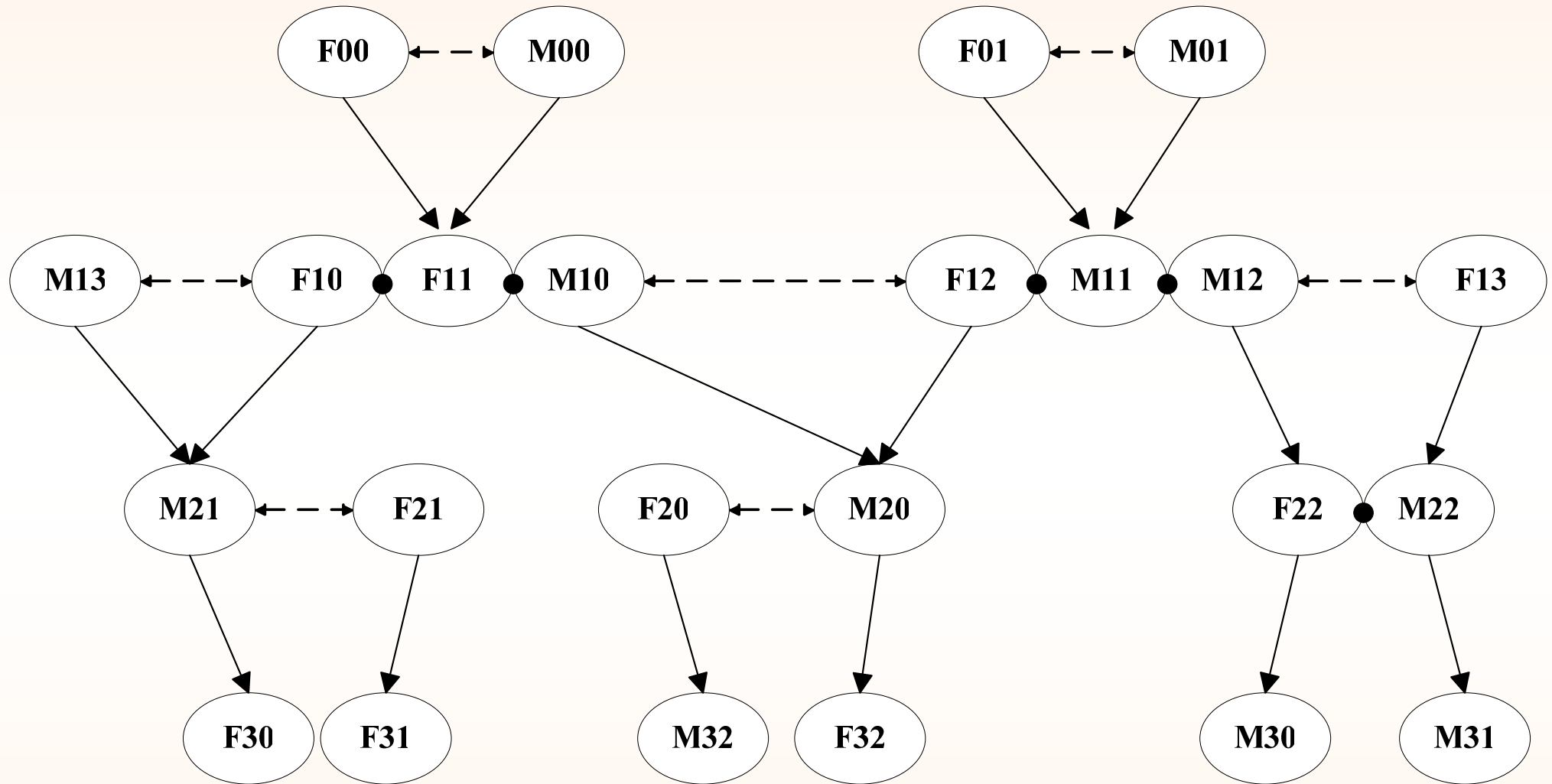
- Wykorzystanie **systemów regułowych** do zadawania zapytań do relacyjnych baz danych
- **Zamiana** ontologii na zbiory reguł
- **Automatyczna** generacja skryptów w języku Jess
- **Odwzorowywanie** pojęć/relacji ontologicznych na schemat relacyjnej bazy danych
- **Bezpośredni** dostęp do bazy danych w trybach wnioskowania wstecz oraz w przód
- Zapytania zadawane **na podstawie pojęć/relacji** dostępnych w ontologii
- **Hybrydowe** metody wnioskowania w realizacji zapytań

Architektura SDL



Przykład

Ontologia w Protégé 4.0.2



Małżonek/Mażonka ← - - - - - → Małżonek/Mażonka

Dziecko ← ————— Rodzic

Rodzeństwo ● Rodzeństwo

maDziecko
 maRodzica
 maBrata
 maSiostrę

Źródła

- http://semanticweb.org/wiki/Main_Page
- <http://www.w3.org/>
- <http://www.ploug.org.pl/>
- <http://ruleml.org>
- <http://www.jessrules.com/>
- <http://protege.stanford.edu/>
- <http://www.ontoprise.de/>
- <http://www.topquadrant.com/>
- <http://owlapi.sourceforge.net/>
- <http://clarkparsia.com/pellet/>
- Wiele innych,
- Opracowania i materiały własne

Pytania???

**DZIĘKUJĘ
ZA
UWAGĘ**



Jaroslaw.Bak@put.poznan.pl