



KOMPUTEROWE PRZETWARZANIE WIEDZY

**Kolekcja prac 2013/2014
pod redakcją Tomasza Kubika**





SPIS TREŚCI



ROZPROSZONE BAZY WIEDZY

B. Kochanowski, M. Krupop

1.1. Ontologie, bazy wiedzy

Tematyka niniejszego projektu obejmuje zapoznanie się z różnymi metodami modelowania rzeczywistości. Znajomość różnych modeli reprezentacji wiedzy będzie stanowiło podstawę do zapoznania się z formalnymi sposobami opisu pewnego fragmentu rzeczywistości. Ten formalny zapis pewnego zbioru wiedzy nosi nazwę ontologii. Pojęcie to jest bardzo silnie związane z bazami wiedzy, które można utożsamiać ze zbiorem ontologii dotyczącej pewnej dziedziny wiedzy. Wspomniane już bazy wiedzy i sposoby zapisu w nich wiedzy będą stanowiły punkt wyjścia do realizacji praktycznej części niniejszego projektu.

W związku z bardzo szybkim rozwojem techniki komputerowej oraz jej stosowania w wielu dziedzinach życia, pojawiła się konieczność komputerowego przetwarzania wiedzy przechowywanej w specjalnych repozytoriach. Aby wiedza ta mogła być przesyłana i przetwarzana istnieje konieczność modelowania wiedzy w taki sposób, aby mogła ona w bardzo prosty i zrozumiały sposób zostać zapisana za pomocą pewnej formalnej reprezentacji. Możliwa jest także sytuacja w której ta sama wiedza może zostać opisana za pomocą kilku różnych modeli. W pracy [?] opisane zostały 2 sposoby modelowania rzeczywistości: za pomocą języka i symboli oraz za pomocą formy graficznej.

W modelowaniu rzeczywistości za pomocą języka, wiedza zapisywana jest za pomocą zdań lub różnego rodzaju wyrażen. Bardzo dobrym przykładem takiego modelowania są instrukcje obsługi urządzeń. Dzięki zawartej tam wiedzy można łatwo dokonać szybkiej diagnozy wadliwie działającego urządzenia. Zaletą tego modelu jest możliwość bardzo szybkiego przekazu wiedzy. Niestety nadaje się on tylko do opisu bardzo prostego modelu rzeczywistości. Bardziej złożony fragment rzeczywistości będzie wymagał większej ilości struktur potrzebnych do jej zapisania. Model ten posiada bardzo istotną wadę. Mianowicie nie jest on modelem uniwersalnym. Oznacza to, że wiedza modelowana w taki sposób będzie zrozumiała jedynie wśród osób jednakowo interpretujących dany język oraz symbole.

1. Rozproszone bazy wiedzy

Drugą z wymienionych metod jest metoda modelowania rzeczywistości za pomocą reprezentacji graficznej. W odniesieniu do językowej metody reprezentacji pozwala ona na zastąpienie opisu słownego (często bardzo skomplikowanego) za pomocą pojedynczego obrazka tzw. piktogramu lub bardziej rozbudowanego schematu. Przykładem gdzie wykorzystano piktogramy jest wspomniana wcześniej instrukcja obsługi. Zamiast skomplikowanego opisu technicznego w reprezentacji językowej dostajemy krótkie instrukcje w postaci piktogramów, jak w danej sytuacji postąpić.

Głównym zadaniem opisanych tutaj metod jest próba modelowania (pewnego opisu, reprezentacji) otaczającego nas świata. Niestety jest to jednak zadanie niezwykle trudne. Wynika to z bardzo dużej złożoności procesów zachodzących w przyrodzie oraz jej skłonności do uwidaczniania niuansów w bardzo wielu płaszczyznach. Z tego właśnie powodu modeluje się jedynie pewne fakty lub fragmenty rzeczywistości tak aby uzyskać z nich jak najwięcej informacji. W tym celu powstało wiele formalnych języków, które dzięki wewnętrznej reprezentacji faktów przekształcają je w użyteczną informację [?].

Konieczność przetwarzania dużej ilości wiedzy wymusiła przechowywanie jej w specjalnych zbiorach zwanych bazami wiedzy oraz wprowadzenie pewnego standardu określającego formalną reprezentację wiedzy w bazie. W informatyce taką reprezentację wiedzy z danej dziedziny określa się terminem ontologia [?].

Ontologia nie jest pojęciem jednoznacznie zdefiniowanym. Przykładem jest tutaj liczba definicji zawartych w artykule [?]. Każda z tych definicji bazuje na innej uzupełnia ją lub porusza całkiem inne aspekty związane z przemyśleniami na temat formalnego sposobu reprezentacji bazy wiedzy. Warto przytoczyć jedną z prostszych definicji, określających ontologię jako „uporządkowany zbiór, składający się ze zbiorów obiektów oraz relacji pomiędzy nimi (binarnych i unarnych)”. Jednym z najważniejszych założeń ontologii jest stworzenie reprezentacji łatwej do zrozumienia zarówno przez ludzi jak i maszyny (komputery).

Tworzenie ontologii wymaga formalnego sposobu zapisu (specjalnych języków) pozwalających na ich opisywanie. Obecnie większość języków służących do opisu ontologii bazuje na tzw. metajęzykach. Najprościej definiując są to języki służące do definiowania innych języków [?].

Jednym z przykładów metajęzyka jest XML. Jest to język bazujący na znacznikach pozwalający na reprezentację danych w postaci specjalnie zdefiniowanej struktury. Obecnie większość języków lub standardów służących do opisu ontologii zbudowanych jest właśnie na bazie XML.

Ontologię można również traktować jako słownik, zasób słów. Słownik ten może być wykorzystany do opisu instancji.

1.2. Cel projektu

Celem projektu jest wykorzystanie dostępnej bazy wiedzy dotyczącej wyspecjalizowanej dziedziny, w tym wypadku informacji opisujących różne style piwa. Baza wiedzy, znaleziona przy pomocy *Swoogle*, została wykorzystana jako zestaw pojęć do opisu ww. dziedziny. Dysponując tym modelem, można wykorzystać

go, importując do własnej bazy wiedzy, która na tej podstawie opisuje własne instancje. Zaletą takiego rozwiązania jest stworzenie architektury rozproszonej. Baza główna może znajdować się w dowolnym miejscu w sieci, wystarczy, że będzie dostępna. Wykorzystując tę bazę, dowolna firma produkująca piwo, będzie w stanie na jej podstawie stworzyć własną bazę, opisującą własne wyroby, tworząc instancje danych stylów piwa. Dysponując niezbędnymi elementami, czyli słownikiem oraz bazami firmowymi, wykorzystującymi ten słownik, można przystąpić do wykorzystania tych rozproszonych informacji w jednym miejscu.

Projekt ten można zrealizować na dwa sposoby. Pierwszym sposobem jest wykorzystanie zapytań w języku SPARQL z poziomu programu i w ten sposób uzyskiwać potrzebne o właściwościach poszczególnych stylów piwa oraz ich instancjach. Podejście to, nie wymaga pobierania tych baz do pamięci komputera, gdyż zapytanie zwraca jedynie wymagane informacje. Można również dokonać agregację poszczególnych grafów reprezentujących bazy wiedzy do jednej struktury grafowej w programie. Do wykonania takiej operacji służy Apache Jena - framework napisany w języku Java. Dopiero na tej zintegrowanej strukturze można odczytywać potrzebne informacje.

1.3. Narzędzia

W celu realizacji wybranego zadania projektowego wykorzystano następujące narzędzia:

- Protege([www.http://protege.stanford.edu/](http://protege.stanford.edu/)) – edytor służący do tworzenia ontologii,
- Apache Jena ([www.http://jena.apache.org/](http://jena.apache.org/)) – jest to API umożliwiające operowanie na bazach wiedzy, ich agregację, wyświetlanie potrzebnych informacji,
- Swoogle ([www.http://swoogle.umbc.edu/](http://swoogle.umbc.edu/)) - wyszukiwarka pozwalająca na znalezienie dokumentów semantycznego Webu na podstawie słów kluczowych; dodatkowo z dokumentów tych wyodrębniane są metadane co służy do określenia relacji między nimi.[?].

1.4. Semantic Beer Base

W ramach projektu powstał program Semantic Beer Base, który umożliwia wczytywanie słownika oraz baz firmowych zawierających instancje różnych stylów piwa. Program, za pomocą Jena łączy poszczególne bazy firmowe w jeden graf, a następnie wyświetla wszystkie te instancje, dodając właściwości reprezentowanego przez nie stylu piwa. Informacje te wyświetlane są za pomocą struktury drzewiastej. Przykład działania aplikacji, wyświetlającej wczytane informacje przedstawione zostały na rys.