

OfficeObjects[®] Ontology Manager

RODAN
DEVELOPMENT

Spis treści

Wstęp.....	3
Geneza produktu	3
Potencjalni odbiorcy systemu	4
Mapy Pojęć – Standard Topic Maps.....	5
<i>Mapy Pojęć a Model Obiektowy i Relacyjny Reprezentacji Informacji</i>	5
<i>“Pojęcie” jako podstawowy budulec Mapy Pojęć</i>	7
<i>Klasy pojęć</i>	7
<i>Nazwy pojęć</i>	8
<i>Atrybuty pojęć</i>	8
<i>Związki pojęć</i>	9
<i>Identyfikacja pojęć</i>	10

Wstęp

OfficeObjects®Ontology Manager – Zarządzanie wiedzą organizacji



OfficeObjects®Ontology Manager jest narzędziem służącym do definiowania, prezentacji, oraz wykorzystania wiedzy o instytucji publicznej lub firmie (ontologii firmy). Wiedza ta może przykładowo obejmować komórki organizacyjne, pracowników, klientów, rodzaje dokumentów, zależności pomiędzy pracownikami (przełożony-podwładny, zastępujący-zastępowany), kompetencje, uprawnienia, słowniki itp.

Wszystkie tego rodzaju pojęcia definiowane są jako elementy mapy pojęć o strukturze zgodnej ze standardem Topic Maps (ISO 13250). Elastyczność wynikająca z tego standardu pozwala na zawarcie w ramach jednej spójnej mapy pojęć takich struktur jak zbiory pojęć, słowniki, słowniki hierarchiczne, hierarchie, sieci semantyczne, grafy, taksonomie, mapowania kategorii, ontologie, tezaury, modele i meta-modele danych. Integracja wiedzy pozwala na jej wykorzystanie praktycznie we wszystkich obszarach funkcjonalnych systemów informatycznych w tym w procesach biznesowych, automatyzacji decyzji, dekretacji zadań, podczas wyszukiwania i katalogowania informacji itp. Istotną cechą narzędzia jest jego zdolność do generowania interfejsów użytkownika tzn. automatyczne dostosowanie interfejsów użytkownika tak, aby końcowi użytkownicy łatwo mogli wprowadzać i modyfikować określone obszary wiedzy.

Geneza produktu

Działalność ludzka w obrębie dowolnej organizacji związana jest z pewną wiedzą niezbędną do efektywnego wykonywania zadań. Każdy nowy pracownik organizacji musi poznać środowisko, w którym będzie pracował tzn. współpracowników, pomieszczenia biurowe, zasoby, zakres zadań, kompetencje, przełożonych, podwładnych, rodzaje dokumentów, spraw, procedury itp. Współczesne systemy informatyczne pozwalają na gromadzenie i wyszukiwanie danych. Zwiększenie efektywności tych systemów możliwe jest poprzez automatyzację procedur w ramach procesów biznesowych dbających o wykonywanie zadań pracowników w odpowiednich czasie i kolejności. Okazuje się, że stworzenie takich procesów nie jest możliwe bez formalizacji wiedzy w sposób czytelny zarówno dla człowieka jak i systemu komputerowego. **OfficeObjects®Ontology Manager** skutecznie realizuje to zadanie pozwalając między innymi na tworzenie merytorycznych, adaptacyjnych procesów pracy oraz inteligentnych rozwiązań.

Potencjalni odbiorcy systemu

OfficeObjects@Ontology Manager skierowany jest do tych instytucji i firm, które chcą podnieść swoją efektywność poprzez zbudowanie inteligentnych i adaptacyjnych systemów informatycznych opartych o wiedzę. Każda działalność ludzka związana jest z pewną wiedzą, dlatego też **OfficeObjects@Ontology Manager** może być wykorzystany zarówno w małych, średnich jak i dużych firmach i jednostkach administracji publicznej. Korzystne jest również zbudowanie bazy wiedzy niezbędnej do przeprowadzenia integracji systemów działających w wielu współpracujących urzędach lub przedsiębiorstwach. Odbiorcami produktu mogą być również producenci oprogramowania pragnący wzbogacić swoje rozwiązania o system reprezentacji wiedzy.

Główne obszary zastosowania

1. Reprezentacja i wykorzystanie wiedzy w dowolnych obszarach systemów informatycznych, a w szczególności w modułach zarządzających procesami biznesowymi, uprawnieniami i konfiguracją.
2. Ułatwienie komunikacji pomiędzy ludźmi i organizacjami poprzez wspólne rozumienie pojęć, redukcję możliwości powstania nieporozumień, jednoznaczną definicję pojęć, integrację wielu punktów widzenia.
3. Zapewnienie współpracy wielu komponentów, modułów, systemów poprzez współdzielenie oraz mapowanie wiedzy.
4. Przejęcie roli słowników systemowych poprzez zapewnienie spójnego, bogatego interfejsu użytkownika, redukcję czasu tworzenia systemu, obsługę wielojęzyczności, wizualizację w formie hierarchii itp.
5. Modyfikowanie systemów w czasie jego działania poprzez parametryzację funkcjonalności w ramach ontologii oraz wykorzystanie generatywnego interfejsu użytkownika (interfejs generowany na podstawie definicji w ontologii).

Główne moduły narzędzia

1. Nawigator Map Pojęć pozwalający na definiowanie i przeglądanie wiedzy za pomocą przeglądarki internetowej,
2. Serwer Map Pojęć pozwalający na dystrybucję wiedzy w ramach wielu systemów i lokalizacji.
3. Silnik Topic Maps oraz API w języku Java pozwalający na efektywne zarządzanie, odczyt i zapis map pojęć z poziomu innych komponentów.
4. Język Skryptowy TMSL oparty o gramatykę javascript pozwalający na łatwe tworzenie zapytań, filtrów, reguł wnioskujących, wyzwalaczy itp.
5. Agent Synchronizacji Informacji z serwerami katalogowymi w oparciu o protokół LDAP.
6. Agent Materializacji Map Pojęć w relacyjnych bazach danych (dla celów raportowych).
7. Zestaw narzędzi administracyjnych, eksportu i importu

Mapy Pojęć – Standard Topic Maps

Umiejętne tworzenie baz wiedzy uzależnione jest od sposobu zapisu informacji w postaci mapy pojęć zgodnie ze standardem Topic Maps.

Angielski termin "topic map" może być tłumaczony jako "mapa tematów". Takie tłumaczenie odzwierciedla pierwotne zastosowanie standardu TopicMaps jako indeksu rozproszonych zasobów. Obecne zastosowania standardu TopicMaps wykraczają jednak poza pierwotne założenia. W kontekście reprezentacji wiedzy bardziej adekwatnym tłumaczeniem jest "mapa pojęć". Termin ten dość dobrze zadomowił się w polskiej terminologii naukowej. Należy mieć jednak na uwadze iż termin "pojęcie" (ang. concept) często utożsamiane jest ze zbiorem obiektów danego rodzaju (np. wszystkie ptaki), natomiast w kontekście map pojęć termin ten oznacza dowolny rzeczywisty lub abstrakcyjny obiekt (ang. subject).

Mapy Pojęć a Model Obiektowy i Relacyjny Reprezentacji Informacji

Na przestrzeni dziesięcioleci rozwoju informatyki pojawiały się coraz to nowe modele reprezentacji informacji. Obecnie powszechnie znane i używane są modele obiektowe i relacyjne. Naturalne jest, że osoby posługujące się na co dzień tymi modelami patrzą na mapy pojęć przez ich pryzmat. Należy jednak mieć świadomość, że mapy pojęć jak i inne modele reprezentacji wiedzy (F-Logic, Frames, Description Logic, RDF), zostały zaprojektowane w celu uwypuklenia specyficznych cech informacji i całkowite ich utożsamianie z modelami obiektowym i relacyjnym jest niewłaściwe. Pomimo tego iż możliwe jest przeniesienie informacji zapisanej w mapie pojęć do modelu obiektowego czy też relacyjnego (jak i odwrotnie), nie można mówić o równości tych modeli. Poniżej opisano cechy charakterystyczne odróżniające mapy pojęć od podejścia obiektowego czy też relacyjnego.

Mapy pojęć zasadniczo służą do reprezentacji wiedzy, w odróżnieniu od modeli przeznaczonych do reprezentacji danych. Wiedza jest to zespół fenomenów umysłowych (czyli tego co zawiera umysł każdego człowieka), które z uzasadnionym prawdopodobieństwem możemy uznać za prawdziwe czyli zgodne z zewnętrzną rzeczywistością. O ile dane mogą zawierać miliony porcji informacji (np. lista transakcji banku) o tyle wiedza jest znacznie mniej liczna i może być zapamiętana i używana przez pojedynczego eksperta w danej dziedzinie (np. lista typów transakcji). Nie ma tutaj ścisłego podziału, dowolna porcja informacji może stać się elementem wiedzy, jeśli tą informację nazwiemy, opiszemy i skojarzymy z innymi elementami wiedzy. Jeśli jakaś porcja informacji jest na tyle istotna, że mówi się o niej w ramach danego środowiska, to informacja ta może stać się elementem wiedzy (np. transakcja na najwyższą kwotę w danym banku). Zasadnicza różnica, jaka się tutaj pojawia pomiędzy mapą pojęć a modelem obiektowym i relacyjnym jest brak z góry narzuconej struktury informacji zapisanej w mapie pojęć. W modelach obiektowym i relacyjnym struktura danych jest ustalona w czasie wstępnej fazy rozwoju systemu informatycznego. W mapach pojęć można dodawać nowy typ

informacji, można tworzyć zupełnie nowe typy związków pomiędzy elementami wiedzy w czasie funkcjonowania systemu. Istotną cechą map pojęć jest, więc ich elastyczność. W modelu obiektowym jak i relacyjnym projektant lub programista określa w czasie budowy systemu, jakie atrybuty może posiadać obiekt reprezentujący informację. W mapach pojęć można dodawać dowolne atrybuty do elementów informacji także w trakcie działania systemu. Sposób łączenia obiektów w modelu obiektowym uzależniony jest od kodu programu, w przypadku map pojęć łączenie elementów wiedzy w związki nie podlega takim ograniczeniom.

Ważnym elementem modeli obiektowych są metody, które wykonują określone funkcje wykorzystując informacje zapisane w obiekcie. Mapy pojęć nie posiadają tego rodzaju mechanizmów aczkolwiek możliwa jest modyfikacja i analiza treści mapy z wykorzystaniem obiektowego API motoru zarządzającego mapami. Funkcje takiego API mają jednak charakter ogólny, a nie specjalizowany dla danego typu informacji przechowywanej w mapie. Odczyt i przetwarzanie wiedzy zawartej w mapie pojęć możliwy jest natomiast poprzez specjalizowane języki np. TMQL (Topic Map Query Language), Tolog. W przypadku modułu Ontology Manager mamy do dyspozycji język TMSL (Topic Map Script Language) stanowiący połączenie języka Javy i elementów języka TMQL. Skrypty w tym języku mogą być pisane i uruchamiane w trakcie działania systemu, co znacznie ułatwi rozwój systemu informatycznego, oraz ekstrakcję informacji w czasie działania systemu.

Model relacyjny w odróżnieniu od większości motorów map pojęć pozwala na przechowywanie olbrzymich zbiorów danych. W większości zastosowań nie ma jednak potrzeby tworzenia map pojęć zawierających więcej niż kilkadziesiąt tysięcy obiektów informacyjnych. Jest to zgodne z założeniem, iż mapa pojęć odzwierciedla deklaratywną wiedzę ekspertów z danej dziedziny w określonym zakresie. Istotna jest natomiast szybka analiza związków pomiędzy pojęciami. W modelu relacyjnym łączenie wielu tabel w zapytaniach SQL skutkuje znacznym spadkiem wydajności. W szczególności w przypadku analizy informacji hierarchicznych kłopotliwe może być skonstruowanie odpowiednich zapytań w języku SQL.

Mapy pojęć traktują związki pomiędzy pojęciami jako elementy konstrukcyjne mapy i odzyskiwanie informacji poprzez analizę złożonych relacji czy też zależności hierarchicznych nie sprawia problemów wydajnościowych. Odpowiednio skonstruowany motor map pojęć zapewnia bardzo wysoką wydajność w porównaniu do relacyjnych baz danych.

Mapy pojęć doskonale sprawdzają się do reprezentacji takich struktur jak:

1. zbiorów pojęć
2. słowników
3. słowników hierarchicznych
4. indeksów rozproszonych zasobów
5. grafów

6. hierarchii
7. sieci semantycznych
8. taksonomii
9. odwzorowania (mapowania) kategorii
10. ontologii
11. tezaurusów
12. modeli i meta modeli danych

“Pojęcie” jako podstawowy budulec Mapy Pojęć

Mapa pojęć jest zbiorem elementów wiedzy odpowiadających rzeczywistym, lub abstrakcyjnym obiektom. Pojedynczy obiekt wiedzy (ang. subject) reprezentowany jest jako pojęcie (topic w standardzie TopicMaps). Pojęcie jest opisane za pomocą zestawów nazw, atrybutów, klas pojęć, identyfikatorów, związków z innymi pojęciami. Elementy te opisane są w dalszej kolejności.

Każde pojęcie może odpowiadać obiektom rzeczywistym, (np. „Amadeusz Mozart”), ich klasom (np. „osoba”), lub nawet obiektom abstrakcyjnych (np. „muzyka”). Ponadto pojęcie może w jednym kontekście odnosić się do klasy obiektów, a w innym stanowić określony obiekt.

Zasadniczym zadaniem inżyniera wiedzy jest odpowiednie odzwierciedlenie świata rzeczywistego w mapie pojęć jako zbioru połączonych i dobrze opisanych pojęć. Inżynier wiedzy poprzez modyfikacje ontologii mapy pojęć w module Ontology Manager może również umożliwić innym użytkownikom systemu kontrolowaną modyfikację mapy.

Klasy pojęć

Zbiór pojęć wchodzących w skład mapy pojęć może być podzielony na podzbiory pojęć należących do określonej klasy. Pozwala to na uporządkowanie pojęć według ściśle określonej kategoryzacji. Należy zwrócić uwagę na to iż jedno pojęcie może należeć do dowolnej liczby klas.

Jeśli pojęcie P należy do klasy K mówimy iż pojęcie P jest instancją klasy K. Często używa się również terminologii: pojęcie P jest typu K.

Pojęcia mogą być rozróżnialne poprzez przyporządkowanie im określonych typów (ang. topic type). Typem pojęcia może być dowolne inne pojęcie znajdujące się w mapie pojęć. Pojęcie może posiadać wiele typów. Pomiedzy pojęciem a pojęciem będącym jego typem istnieje więc swoista relacja. Relacja taka nazywana jest często relacją pomiędzy klasą a instancją. Pojęcie będące typem jest klasą, a pojęcie posiadające typ jest instancją tej klasy. Przykładowo „Amadeusz Mozart” jest pojęciem posiadającym przyporządkowany typ „kompozytor”. Pojęcie „kompozytor” jest w tym wypadku klasą, a pojęcie „Amadeusz Mozart” jest instancją tej klasy. Jednocześnie zauważmy, że pojęcie „kompozytor” mimo tego iż samo jest klasą może być instancją innych klas np. klasa

„profesja”, lub „rodzaj udziałów w tantiemach”.

Mapy pojęć umożliwiają również dziedziczenie klas. Przykładowo klasy pojęć „mężczyzna”, „kobieta” mogą dziedziczyć z klasy „osoba”. Klasa może posiadać wiele nad-klas (klas bazowych). Jeżeli klasa X dziedziczy z klasy Y to pojęcie należące do klasy X posiada cechy zarówno klasy X, jak i Y.

Ponieważ zarówno relacja klasa – instancja jak i klasa bazowa – klasa potomna są związkami pojęć często osoby nie posiadające bogatej praktyki myślą te dwie relacje. Relacja dziedziczenia jest w rzeczywistości sumowaniem podzbiorów pojęć należących do klas potomnych. Relacja klasa - instancja dotyczy przynależności jednego pojęcia do określonego podzbioru pojęć.

Nazwy pojęć

Określone elementy rzeczywistości mogą posiadać wiele nazw – w szczególności nazwy w różnych językach. Standard Topic Maps wychodzi naprzeciw tego rodzaju wymaganiom - pozwala na przypisywanie do jednego pojęcia wielu nazw (ang. topic name). Nazwa stanowi ciąg znaków. Elementem pozwalającym odróżnić nazwy od siebie jest tzw. zakres (ang. scope). Każda z nazw może mieć przyporządkowany zakres określający w jakim kontekście powinna ona być używana. Zakres definiowany jest jako zbiór pojęć, dla przykładu może on składać się z pojęć: „język polski”, „domyślna”. Nazwa posiadająca tak zdefiniowany zakres powinna być używana jeśli osoba, korzystająca z wiedzy zawartej w mapie pojęć, posługuje się językiem polskim.

Standard Topic Maps nie nakłada żadnych ograniczeń na liczbę nazw danego pojęcia. Nie ma więc problemów z nadaniem kolejnych nazw dla pojęcia w czasie działania systemu. Inżynier wiedzy powinien dbać jednak o to, aby możliwe było rozróżnienie nazw za pomocą zakresu widoczności. W szczególności jeśli pojęcie posiada kilka nazw w tym samym języku to tylko jedna z nich powinna posiadać pojęcie „domyślna” w określającym ją zakresie. W ten sposób można uniknąć sytuacji, w której komponent wizualizacji mapy pojęć zamiennie stosowałby różne nazwy do prezentacji tego samego pojęcia.

Standard Topic Maps definiuje również warianty nazwy. Każda nazwa pojęcia może zawierać wiele wariantów opisywanych przez określone właściwości. Przykładowo określona nazwa w języku polskim może mieć wersje w liczbie mnogiej. W obecnej wersji moduł Ontology Manager nie wspiera wariantów nazw, które mogą być jednak zrealizowane za pomocą podstawowych nazw o określonych zakresach.

Atrybuty pojęć

Standard Topic Maps stworzony został dla celów indeksacji rozproszonych zasobów. Nieodzownym elementem mapy pojęć jest więc tzw. wystąpienie (ang. occurrence). Wystąpienie wskazuje na zasób

związany z pojęciem. Najczęściej zasób identyfikowany jest przez adres w formie URI (Uniform Resource Identifier), ale dopuszczalne są inne formy adresowania. Pojęcie może posiadać wiele wystąpień, które rozróżnialne są za pomocą typu oraz zakresu wystąpienia. Typ wystąpienia to dowolne pojęcie znajdujące się w mapie pojęć określające jakiego rodzaju zawartość dostępna jest we wskazanym zasobie np. „rysunek”, „dokumentacja” itp. Zakres wystąpienia określa kontekst w jakim dane wystąpienie może być stosowane (np. język, w którym napisano zasób wskazywany przez wystąpienie). Podobnie jak to jest w przypadku nazw nie ma ograniczeń co do ilości atrybutów pojęcia.

Standard dopuszcza również na stosowanie tzw. bezpośrednich wystąpień (ang. in-line occurrence). Wystąpienia tego rodzaju zapisane są bezpośrednio w mapie pojęć w postaci ciągu znaków. Mogą to być np. definicje, opisy pojęć itp. Mechanizm ten można wykorzystać do zapisu metadanych związanych z pojęciami.

W module **OfficeObjects@Ontology Manager** bezpośrednie wystąpienia jak i wystąpienia w postaci odniesień do zasobów zewnętrznych wskazywanych przez określone URI nazywane są atrybutami pojęcia. Jest to termin bliższy modelom reprezentacji wiedzy i łatwiejszy do przyswojenia przez użytkownika.

W odróżnieniu od modeli obiektowych i relacyjnych nie ma ścisłego związku pomiędzy klasą pojęć a listą atrybutów i nazw instancji tej klasy – dwa pojęcia należące do tej samej klasy mogą posiadać odmienne listy typów atrybutów. Taka elastyczność choć czasami niezwykle pomocna stwarza duże problemy w organizacji specjalizowanych interfejsów użytkownika. Dlatego też OfficeObjects@Ontology Manager wprowadza mechanizm schematów klas pojęć pozwalający zdefiniować jakie nazwy i atrybuty są dozwolone dla pojęć należących do określonych klas. Schematy te używane są na poziomie Nawigatora Map Pojęć (graficzny interfejs użytkownika służący do edycji mapy pojęć). Na poziomie motoru map pojęć programista nie ma ograniczeń co do listy atrybutów i nazw danego pojęcia bez względu na klasy, do których pojęcie należy.

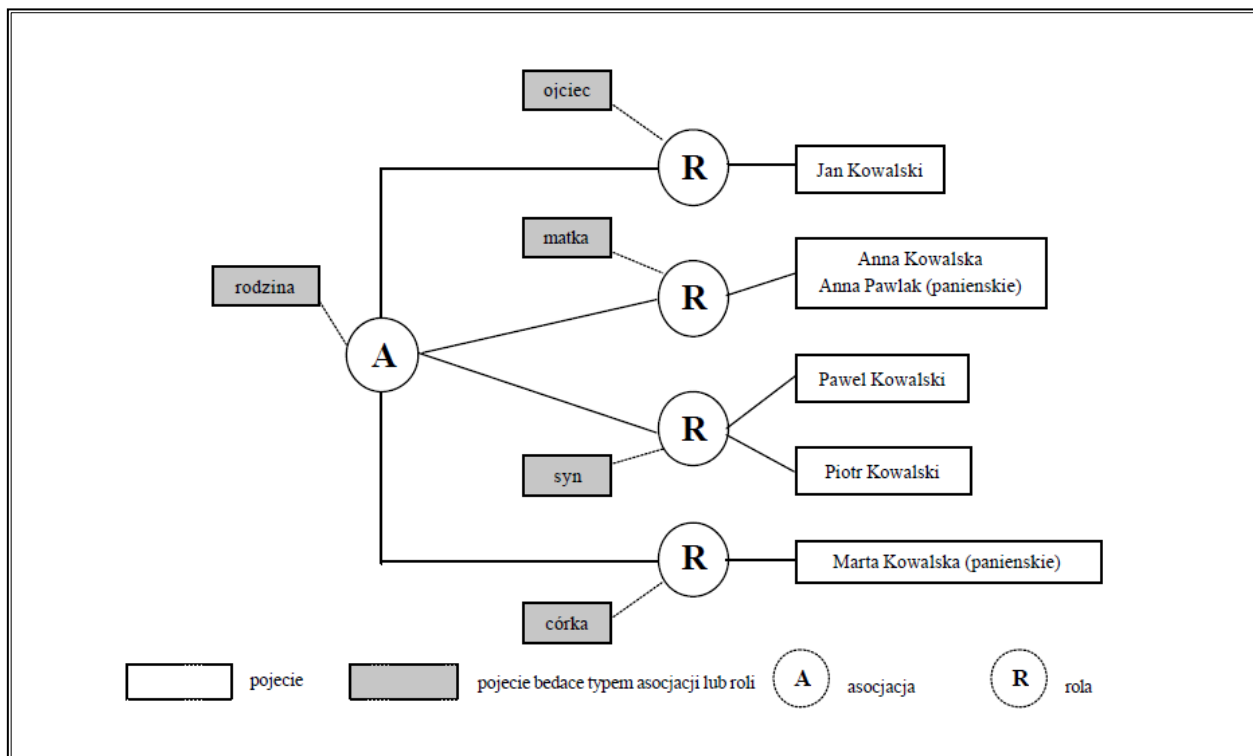
Związki pojęć

Mapa pojęć byłaby zwykłym słownikiem gdyby nie asocjacje (ang. association) pomiędzy pojęciami. Asocjacje mogą wiązać wiele pojęć. Każda asocjacja posiada typ określony za pomocą dowolnego pojęcia. Pojęcia wchodzące w skład asocjacji (ang. association members) pełnią określone role w tej asocjacji (ang. topic plays role). Typ roli to dowolne pojęcie w mapie pojęć.

Każdy związek pojęć posiada więc następujące cechy:

- typ związku określony przez dowolne pojęcie
- zakres (zestaw pojęć, które określają zakres użycia związku)
- wiąże dowolną liczbę pojęć

- każde pojęcie w związku gra określoną rolę, tak więc związek składa się z zestawu ról pojęć.
- typ roli pojęcia w związku określony jest przez dowolne pojęcie.



Rysunek 1 Przykładowy związek pojęć

Rysunek 1 ilustruje przykładowy związek pojęć. Związek ten jest typu "rodzina" i posiada następujące role "ojciec", "matka", "syn", "córka". Związek można przedstawić w postaci poniższej tabeli.

Rodzina			
Ojciec	Matka	Syn	Córka
Jan Kowalski	Joanna Kowalska	Michał Kowalski	Julia Kowalska Anna Kowalska

Tabela 1: Przykładowy związek pojęć.

Jak widać w ostatniej komórce tabeli mamy dwie wartości. Model związków map pojęć nie odpowiada więc ściśle modelowi relacyjnych baz danych, jest od niego znacznie szerszy. Za pomocą związków w mapach pojęć można przedstawić dowolny wiersz tabeli relacyjnej bazy danych. Odzworowanie w drugą stronę może wymagać zastosowania wielu tabel bazy danych.

Identyfikacja pojęć

Kolejnym istotnym elementem map pojęć jest mechanizm identyfikacji pojęć. Mechanizm taki powinien zapewniać możliwość łączenia map pojęć wdrożonych w odrębnych systemach podlegających

integracji. Mechanizm ten również służy do zapewnienia niepowtarzalności pojęć. Pojęcia identyfikowane są za pomocą URI. Istnieją dwa rodzaje identyfikatorów. Pierwszym z nich jest tzw. lokalizacja pojęcia (ang. subject locator). Ten sposób identyfikacji używany jest w sytuacjach gdy pojęcie odnosi się do obiektu będącego zasobem sieciowym wskazywanym przez URI. Jeśli przykładowo pojęciem jest "Strona Internetowa NASA" to do identyfikacji używamy adresu tej strony. Jeśli natomiast pojęciem byłoby "NASA" to wtedy używamy drugiego sposobu identyfikacji za pomocą tzw. wskazania obiektu (ang. subject indicator). W takiej sytuacji URI nie wskazuje bezpośrednio na obiekt do którego odnosi się pojęcie, lecz wskazuje jakiś zasób opisujący ten obiekt.

Dwa pojęcia identyfikowane przez tą samą lokalizację zasobu lub to samo wskazanie obiektu traktowane są jako to samo pojęcie i jest łączone przez motor przetwarzające mapy pojęć.

URI identyfikujące pojęcie często jest wykorzystywane w kodzie oprogramowania dzięki czemu program może odwoływać się do ściśle określonego pojęcia w mapie.