

ASPEKTY IMPLEMENTACJI INTELIGENTNEJ INFRASTRUKTURY INFORMACJI PRZESTRZENNEJ*

IMPLEMENTATION ASPECTS OF AN INTELLIGENT SPATIAL INFORMATION INFRASTRUCTURE

Adam Iwaniak^{1,3}, Jaromar Łukowicz², Marek Strzelecki³, Iwona Kaczmarek¹

¹Uniwersytet Przyrodniczy we Wrocławiu

²STRUKTURA, Planowanie Przestrzenne, GIS

³Kon-Dor GIS Consulting

Słowa kluczowe: ontologie, interoperacyjność semantyczna, inteligentna infrastruktura informacji przestrzennej, RDF, OWL

Keywords: ontologies, semantic interoperability, intelligent spatial information infrastructure, RDF, OWL

Wstęp

Artykuł przedstawia koncepcję systemu informatycznego, łączącego klasyczne oprogramowanie GIS z technologią sieci semantycznych. Polega ona na integracji danych przestrzennych, jak również informacji i wiedzy dotyczącej przestrzeni, które określają zakres działania systemu. W tym celu wykorzystywane są tezaury, ontologie oraz służące do ich opisu języki RDF i OWL. Zaproponowane podejście pozwala budować inteligentne systemy do analiz i wspomagania decyzji w dziedzinie gospodarki przestrzennej, obejmującej zagadnienia planowania przestrzennego, lokalizacji inwestycji, gospodarki nieruchomościami, ochrony środowiska oraz zarządzania infrastrukturą techniczną i drogową. Opisane rozwiązania stanowią próbę praktycznej implementacji idei inteligentnej infrastruktury informacji przestrzennej.

Sieci semantyczne jako narzędzie budowy globalnej bazy WWW

Podstawą dostępu do informacji w Internecie jest sieć WWW. Jej sukces wynika z elastyczności i doskonałego przystosowania do otwartej architektury infrastruktury Internetu. Rozwiązuje potrzebę powszechnego dostępu wszelkich podmiotów do jego zasobów, zarówno w roli dostawców treści, jak i ich odbiorców. Cechami sieci WWW są:

- powszechność – każdy użytkownik WWW może być zarówno odbiorcą, jak i dostawcą treści,

* Artykuł sfinansowany przez Narodowe Centrum Nauki ze środków przeznaczonych na naukę.

- brak hierarchii zasobów, centralnego zarządzania architekturą oraz strukturą zasobów,
- otwartość i swoboda implementacji zasobów – każdy dostawca może swobodnie tworzyć i rozbudowywać własne zasoby, umieszczać je w dowolnych węzłach sieci Internet, a struktura wewnętrzna zasobów stanowi zakres decyzji poszczególnych dostawców,
- elastyczność – architekturę WWW można łączyć z nowymi technologiami informatycznymi, zwiększając jej funkcjonalność: dynamika (bieżąca aktualizacja treści) oraz interaktywność (wpływ na udostępnianą treść, aplikacje sieciowe),
- skalowalność – sieć WWW, poza ograniczeniami fizycznymi, nie jest ograniczona z punktu widzenia architektury i może obsługiwać Nielimitowaną liczbę domen, udostępnianych w nich zasobów oraz nieograniczoną wielkość treści zawartych w tych zasobach.

Fundamentalną koncepcją architektury i technologii WWW, zapewniającą powyższe zalety, jest nieustrukturalizowana forma i dystrybucja zasobów. Zapewniają ją następujące elementy:

- sieć Internet, infrastruktura stanowiąca fizyczny nośnik zasobów, ze standardami (TCP/IP) oraz usługami (DNS),
- Uniform Resource Identifier (URI), standard zapewniający jednoznaczność lokalizację i identyfikację zasobu, bez konieczności hierarchizacji struktury WWW,
- Hypertext Markup Language (HTML), standard zapisu dokumentów w WWW,
- Hypertext Transfer Protocol (HTTP), protokół stanowiący standard udostępniania zasobów w Internecie.

Elastyczność i skalowalność WWW pozwala poszerzać obszary jej zastosowań i funkcjonalność. Sieć ta gromadzi i udostępnia niespotykane dotąd w historii rozmiary treści, zawierające niezliczoną liczbę danych i informacji. Ze względu na prezentacyjną orientację HTML, którego głównym zadaniem jest formatowanie wizualne zawartości stron internetowych, zasoby te są czytelne niemal wyłącznie dla człowieka (ang. *humanreadable*). Poza zaawansowanymi metodami analizy tekstu w języku naturalnym, które przynoszą ograniczone rezultaty, zasoby te nie nadają się do automatycznego przetwarzania przez maszyny liczbowe. Zasobów WWW nie można więc traktować jako użytecznego źródła danych dla procesów informatycznych, analogicznie jak w przypadku różnego typu baz danych.

Jak zatem wykorzystać zasoby WWW, aby ich bogata zawartość, stała się ogólnie dostępną bazą danych? Jednym ze sposobów jest zdefiniowanie wyspecjalizowanych usług, nastawionych na realizację ściśle określonych funkcji oraz dostarczanie danych w jednoznacznie zdefiniowanych schematach. Taka koncepcja, w odniesieniu do danych geograficznych, realizowana jest w ramach infrastruktury danych przestrzennych (ang. *Spatial Data Infrastructure – SDI*), która wykorzystując technologię WWW, definiuje wyspecjalizowane usługi z własnymi protokołami oraz dostarcza danych w sztywno zdefiniowanych schematach. Korzystanie z tej infrastruktury pozwala na automatyzację wykorzystania udostępnianych zasobów danych przestrzennych w celu uzyskania informacji o określonych zjawiskach i obiektach przestrzennych.

Odmienne podejście prezentuje twórca WWW, Timothy Berners-Lee. Jego idea polega na przekształceniu zasobów WWW w globalną bazę wiedzy, bez wprowadzania sztywnych standardów ograniczających uniwersalność usług. Semantic Web, umożliwił ma automatyczne (komputerowe) przetwarzanie zasobów internetowych WWW, poprzez nadanie im

„znaczenia”. Opis zawartości WWW, który nadawałby się do interpretacji przez powszechnie dostępne oprogramowanie, uczyniłby treści gromadzone w sieci „zrozumiałe dla komputerów” (ang. *machinereadable*).

Semantic Web jest środkiem na wzbogacenie zasobów WWW o sformalizowaną warstwę opisu (metadane), bez zmiany architektury samej usługi WWW. Metadane nadają znaczenie zasobom poprzez odwołanie się do modeli pojęciowych, prezentujących relacje pomiędzy pojęciami: logiczne, semantyczne, mereologiczne. W tym celu Semantic Web korzysta z takich technologii jak:

1. **Linked data** – metoda nadawania danym struktury poprzez ich wzajemne powiązanie. Architektura linked data ma formę trypletów, które wiążą jeden zasób (podmiot) z drugim (przedmiot) za pośrednictwem właściwości (predykatu). Linked data korzysta ze standardów stworzonych dla WWW (URI i HTTP) oraz standardu RDF (*Resource Description Framework*), który leży u podstaw wszystkich technologii sieci semantycznych.
2. **Ontologie** – stanowiące formalną reprezentację wiedzy. Według Goczyła (2011) ontologia stanowi formalny opis pojęć występujących w danej dziedzinie. Najpowszechniejsza definicja ontologii pochodzi od Grubera (1993) i brzmi: *Ontologia jest formalną, jawną specyfikacją wspólnej konceptualizacji*. Szerzej, ontologie definiują hierarchię klas obiektów oraz hierarchię ich właściwości (ang. *properties*). W najprostszej formie stanowią one proste słowniki lub bardziej rozbudowane o relacje semantyczne tezaury lub też struktury w postaci taksonomii, jak ontologie zdefiniowane w RDF Schema lub OWL Lite. Bardziej zaawansowane są ontologie wykorzystujące pełną ekspresję systemu formalnego logiki opisowej (ang. *Description Logic* – DL). Budują one hierarchię klas wraz z definiowaniem restrykcji na właściwościach. Pozwala to na zasycenie bogatszej informacją niż tylko dzięki opisowi relacji pomiędzy różnymi zasobami. Ontologie DL zwykle zawierają dwa komponenty. Po pierwsze stanowią definicję modelu logicznego (taksonomia obiektów wraz z właściwościami). Ten typ określany jest, jako TBox (ang. *Terminological Box*). Po drugie jako zbiór obiektów – ABox (ang. *Assertional Box*), zwanych też indywiduami, przypisanych do poszczególnych klas, zdefiniowanych w TBox. Z punktu widzenia budowy dokumentu ontologii, TBox i ABox, tworząc dwa odrębne komponenty, mogą zarówno występować w jednym pliku, jak i osobno w odrębnych dokumentach powiązanych importem. Ontologia DL nie musi zawierać obu części, często zawiera tylko terminologiczną. Ontologie oparte na logice opisowej (zapisywane np. w OWL DL), pozwalają na wiele scenariuszy integracji danych. Można tutaj stosować różne metody łączenia/integracji ontologii (ang. *ontologymerging*), porównywania lub mapowania ontologii (ang. *ontologymatching*, *ontologyalignment*, *ontologymapping*). Techniki te tworzą ujednoczone drzewa klas oraz właściwości. Dzięki temu opis obiektów reprezentowanych w różnych ontologiach zostaje wzbogacony (ang. *enrichement*) zarówno w zakresie klasyfikacji, jak i w zakresie zestawu cech dla poszczególnych obiektów.

Wszystkie powyższe elementy korzystają z rozwiązań i standardów już obecnie stosowanych w technologiach internetowych. Zasoby semantyczne są identyfikowane przez globalny identyfikator URI, co umożliwia ich integrację. Do zapisu RDF i ontologii można wykorzystywać różne języki. Najczęściej wykorzystywane oparte są na XML (RDF/XML, OWL/XML). Przyjęcie dla Semantic Web standardów W3C oraz elastyczność architektury WWW pozwala na powiązanie powyższych elementów z dokumentami HTML/XHTML. Stosuje się tu między innymi następujące specyfikacje: RDFa, GRDDL, Microformats.

Technologie semantyczne, umożliwiają eksplorację zasobów oraz odkrywanie nowych faktów, dzięki narzędziom zapytań, wnioskowania oraz analizy logicznej:

- 1) języki zapytań – tym czym dla baz relacyjnych jest SQL, tym dla grafów RDF jest SPARQL; SPARQL ma składnię podobną do SQL, przy czym warunki ograniczające zwracaną kolekcję obiektów odnoszą się nie do wartości kolumn w tabelach, ale do wartości określonych predykatów (relacji) w grafie;
- 2) systemy wnioskowania (ang. *reasoners*) – są adresowane do ontologii; pozwalają odkrywać nowe fakty, takie jak przynależność obiektów do określonych klas zdefiniowanych przy pomocy ograniczeń na właściwościach (ang. *restrictions*);
- 3) języki regułowe – technologie umożliwiające wnioskowanie, wykorzystujące logikę predykatów pierwszego rzędu (ang. *First Order Logic*); w technologiach sieci semantycznych wykorzystywany jest język SWRL, łączący konstrukcje języka RuleML ze składnią OWL.

Inteligentna Infrastruktura Informacji Przestrzennej

Wraz z popularyzacją idei Semantic Web, rozpoczęto prace nad zaadaptowaniem rozwiązań semantycznych w budowie SDI i GIS. Główne obszary badawcze dotyczą:

- zwiększenia funkcjonalności wyszukiwania danych przestrzennych (Klien i in., 2004; Lieberman, 2006; Lutz, Klien, 2006; Kubik, 2010; Olfat i in., 2012),
- integracji danych przestrzennych (Wache i in., 2001; Iwaniak i in., 2011; Kaczmarek i in., 2011),
- budowy oraz orkiestracji semantycznych usług danych przestrzennych (Andrei i in., 2008; Janowicz, 2010; Lia i in., 2011),
- budowy systemów GIS wykorzystujących technologie sieci semantycznych między innymi dla potrzeb wspomagania podejmowania decyzji (Fonseca i in., 2002; Łukowicz i in., 2012; Jelokhani-Niaraki, Malczewski, 2012).

Podstawowym powodem dla którego wykorzystywane są sieci semantyczne w SDI jest jej heterogeniczny charakter. Heterogeniczność rozważana jest na dwóch poziomach: danych i semantyki (Donato, Pasquale, 2010). Heterogeniczność danych odnosi się do zagadnień związanych z formatem zapisu i typem danych, podczas gdy heterogeniczność semantyczna dotyczy znaczenia i rozumienia informacji. Ten typ heterogeniczności można rozpatrywać w dwóch kategoriach:

- 1) heterogeniczność odnosząca się do nazewnictwa, występująca w sytuacji gdy odmienne od siebie wyrażenia dotyczą tego samego pojęcia,
- 2) heterogeniczność pojęciowa, związana z różną interpretacją tego samego zjawiska czy pojęcia.

Wykorzystanie doświadczeń w obszarze technologii sieci semantycznych jest jedną z propozycji zmierzających do osiągnięcia interoperacyjności semantycznej. Technologie sieci semantycznych pozwalają bowiem na formalny zapis semantyki przekazywanej informacji, która jest czytelna dla maszyn, zatem może być dzielona i „rozumiana” przez wiele systemów informatycznych. Problem właściwego „rozumienia” przesyłanej informacji nie był tak znaczący w sytuacji, gdy systemy GIS miały charakter scentralizowany, a dane, jak i metody ich przetwarzania, gromadzone były lokalnie. W chwili obecnej, kiedy istnieją możliwości pozyskania danych z wielu, różnych, nie powiązanych ze sobą źródeł, semantyka danych nabiera znaczenia.

Obecne podejścia, zmierzające do osiągnięcia interoperacyjności semantycznej w obszarze GIS, oparte są na tworzeniu ontologii. Prace nad wykorzystaniem ontologii w GIS trwają od wielu lat i wciąż stanowią jeden z najmniej odkrytych i obiecujących obszarów badań. Donato i Pasquale (2010) wyróżniają dwa podejścia w zakresie wykorzystania ontologii w GIS:

- 1) podejście filozoficzne – związane z problematyką tworzenia ontologii wyższego poziomu dla geoinformacji, próbą odmiennego kategoryzowania pojęć i innego postrzegania rzeczywistości przez wielu ludzi i różne kultury,
- 2) podejście technologiczne – związane z dziedziną inżynierii wiedzy, traktującą wykorzystanie ontologii, jako narzędzi służących do implementacji konkretnych rozwiązań.

Jednym z podstawowych wyzwań w wykorzystaniu ontologii w GIS jest integracja samych ontologii wykorzystywanych między innymi do opisu danych przestrzennych. W tym zakresie stosuje się wiele podejść, między innymi polegające na tworzeniu jedynie ontologii wyższego rzędu, budowie dla każdego zbioru informacyjnego odrębnych ontologii czy też podejście mieszane, polegające na tworzeniu ontologii dziedzinowych, które wykorzystują pojęcia i relacje z ontologii wysokiego poziomu.

Istotną rolę w tym procesie odgrywają miary podobieństwa semantycznego (podobieństwo znaczenia), dostarczające mechanizmów porównywania terminów w ontologiach i oceny ich poprawności. Efektywność procesu łączenia różnych ontologii stanowi element kluczowy dla integracji danych i osiągnięcia interoperacyjności semantycznej w SDI.

Dla dodania znaczenia do publikowanych usług i danych przestrzennych Janowicz, i inni (2009; 2010) zaproponowali stworzenie semantycznej warstwy dla geoinformacyjnych usług sieciowych zgodnych ze standardami OGC. Infrastrukturę informacji przestrzennej wykorzystującą technologie sieci semantycznych, a w szczególności ontologie, stosowane do opisu danych przestrzennych i usług sieciowych, kontekstowego wyszukiwania, integracji i wnioskowania – autorzy proponują nazwać **Inteligentną Infrastrukturą Informacji Przestrzennej**.

W następnym rozdziale przedstawiono koncepcję rozwiązania wykorzystującego technologie sieci semantycznych dla potrzeb wspomagania podejmowania decyzji oraz opis środowiska programowego wykorzystywanego do jego budowy. Opisywana koncepcja oraz środowisko zostały wykorzystane do opracowania systemu (Łukowicz i in., 2012).

Koncepcja systemu

Uwarunkowania technologiczne

Projektowane środowisko ma za zadanie wspomóc przeprowadzanie złożonych analiz i podejmowanie decyzji wymagających pozyskania informacji z licznych i różnorodnych zasobów danych przestrzennych i opisowych. W pierwszym rzędzie obejmują one dane udostępniane w SDI w postaci serwisów WMS, WFS i WCS, opisane na poziomie serii i zbiorów danych przy pomocy metadanych. W drugiej kolejności dotyczą one danych źródłowych, które mogą być udostępniane w sieci WWW z wykorzystaniem technologii RDF, RDFa, OWL, GRDDL lub Microformats.

Poszczególne narzędzia Semantic Web mogą być wykorzystane na różnych etapach procesów obróbki danych przestrzennych. Po pierwsze, pozwalają one dodać wzbogacony opis

danych, uzupełniając płytki profil metadanych SDI, metadanymi dostarczanymi w formie RDF. Profile metadanych SDI są ograniczone do opisu usług, serii danych oraz zbiorów danych. RDF nie ma tych ograniczeń i pozwala na wprowadzenie dodatkowego opisu zarówno dla obiektów, jak i ich atrybutów. Technicznie do tego celu można wykorzystać specyfikację GRDDL, pozwalającą na zaszywanie opisu w pliku GML lub też można zastosować zewnętrzny opis w plikach RDF z referencjami do identyfikatorów obiektów w SDI. RDF pozwala również określić relacje pomiędzy obiektami. Powiązane właściwościami obiekty tworzą graf RDF, w którym właściwości stanowią krawędzie grafu, a obiekty – węzły (podmioty i przedmioty relacji). Struktura metadanych RDF może być zdefiniowana w ontologiach.

Projektowane rozwiązanie wykorzystuje ontologie również do definiowania problemów decyzyjnych. Ontologie wykorzystujące logikę opisową mogą być doskonałym narzędziem do zapisu bazy wiedzy, zawierającej kryteria decyzyjne dla różnych zagadnień (analiz inwentaryzacyjnych, rozstrzygnięć projektowych, decyzji administracyjnych). Również w tym przypadku środkiem dla zapisu kryteriów decyzyjnych będą restrykcje na właściwościach, definiujące klasy obiektów.

Przeznaczenie i funkcjonalność

Projektowane rozwiązanie może być adresowane do wielu odbiorców. Może być przykładowo wykorzystywane zarówno przez administrację publiczną, przez podmioty gospodarcze, ale również przez obywateli, którzy występują w roli indywidualnych inwestorów. Powinno ono zawierać komponenty, które będą obsługiwać kolejne czynności procesu analizy i podejmowania decyzji oraz pozwolą efektywnie wykorzystać i zaprezentować rezultaty tego procesu. Tego typu zadanie obejmuje następujące sekwencje:

- 1) określenie problemu analitycznego i tworzenie bazy wiedzy służącej jego rozwiązaniu,
- 2) wyszukiwanie odpowiednich do problemu zasobów,
- 3) import surowych danych i interpretacja informacji na temat gromadzonych w nich obiektach,
- 4) integracja danych, polegająca na mapowaniu dostępnych obiektów i ich atrybutów,
- 5) przeprowadzenie procesu wnioskowania wykorzystującego bazę wiedzy oraz zintegrowane dane zewnętrzne i własne przy wykorzystaniu silników wnioskowania,
- 6) transformacja wyników analiz do modelu GIS ze wzbogaconym zestawem atrybutów i opisem wynikowej klasyfikacji obiektów,
- 7) wizualizacja wyników analiz, będących warstwami GIS, w postaci map tematycznych.

Projektowany system budowany jest poprzez powiązanie narzędzi GIS, operujących na zasobach SDI oraz na własnych danych przestrzennych, z narzędziami technologii sieci semantycznych. Architektura systemu zawiera moduły do edycji i obróbki danych przestrzennych oraz definiowania ontologii i prowadzenia wnioskowania. Pomiędzy tymi komponentami jest transformacja z modelu danych SDI, z dedykowanym dla niego schematem aplikacyjnym, do modelu semantycznego, zdefiniowanego w komponentach terminologicznych TBox, czyli do postaci ABox i w drugim kierunku z modelu semantycznego do formatu danych we właściwym schemacie aplikacyjnym. Architektura systemu jest skonstruowana zgodnie z poniższymi założeniami:

1. W ramach projektowanego systemu baza wiedzy będzie miała charakter ontologii, w której, wykorzystując system formalny DescriptionLogic, zostaną określone warunki i wymogi, według których będą badane zjawiska przestrzenne. Warunki będą sformuło-

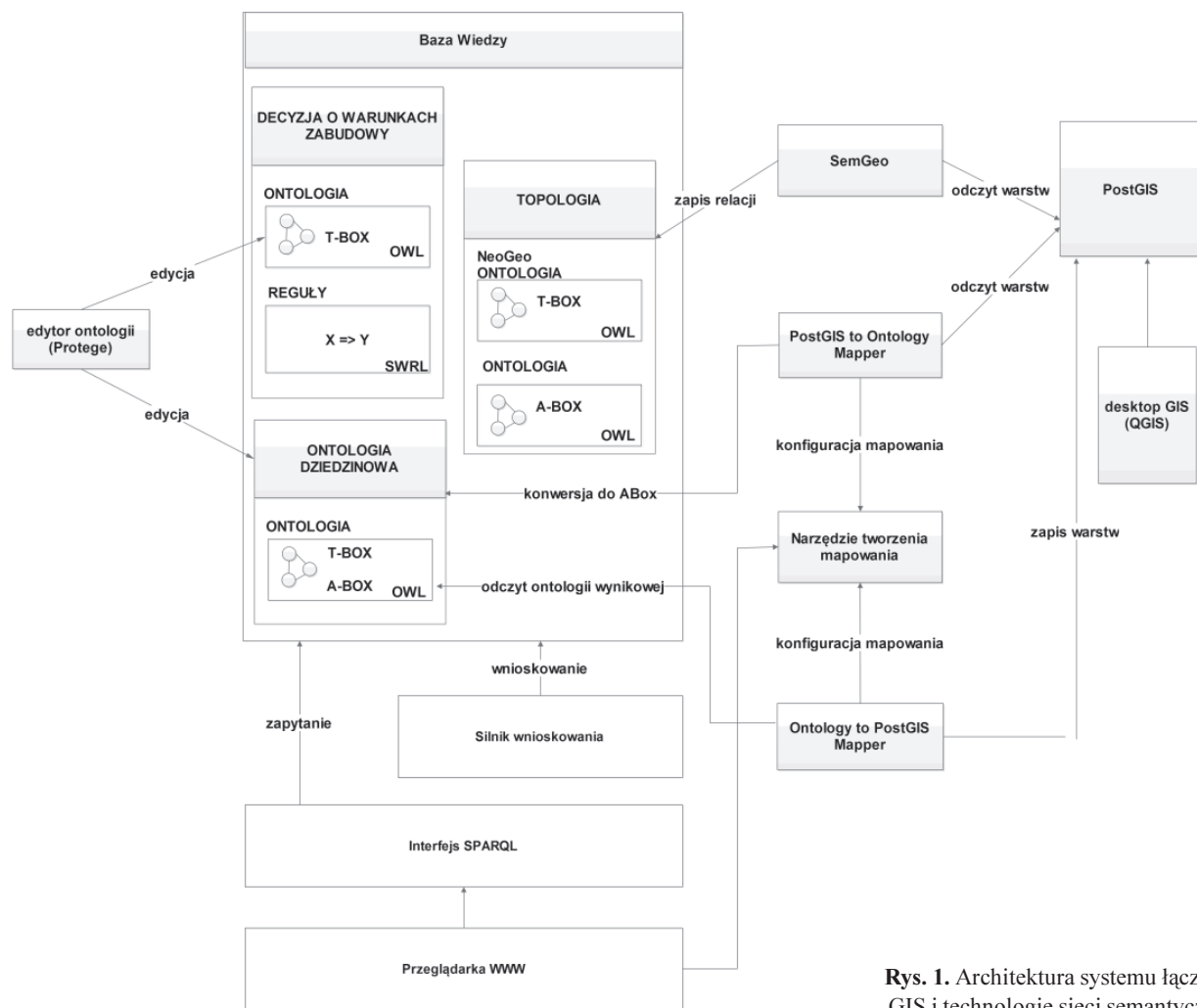
wane poprzez zdefiniowanie hierarchii klas, zawierającej klasy rozwiązań badanych zjawisk, hierarchii właściwości opisującej cechy podlegające analizie oraz poprzez zgodny z systemem DescriptionLogic, system ograniczeń na tych właściwościach dla poszczególnych klas wynikowych.

2. Wyszukiwanie zasobów będzie dokonywane poprzez interpretację dostarczanych przez dostawców metadanych. Metadane te będą przekształcane do ontologii łączonej z ontologiami użytkownika i w drodze wnioskowania wskazywane będą zasoby spełniające wymagania w kontekście prowadzonych analiz.
3. Interpretacja danych pozyskiwanych w SDI oraz WWW odbywać się będzie poprzez mapowanie modeli logicznych SDI na ontologie zgodne z opisem badanej problematyki, stanowiące tzw. TBox modelu, według którego prowadzona jest analiza.
4. Import danych będzie polegał na transformacji obiektów z zasobów SDI oraz WWW do grafu RDF, czyli ABox, przy równoczesnym wzbogaceniu go o opis relacji topologicznych pomiędzy obiektami w ramach poszczególnych zasobów i pomiędzy nimi. ABox będzie zgodny z przygotowanymi ontologiami typu TBox, a import obiektów z atrybutami będzie uwzględniał mapowanie modelu danych SDI na model analityczny zdefiniowany w bazie wiedzy.
5. Wyniki wnioskowania zostaną zachowane w ontologii wynikowej, zawierającej obiekty o nowej klasyfikacji (TBox + ABox). Wnioskowanie wzbogaci więc opis źródłowych obiektów o nowe atrybuty i przypisze je nowym klasom, zdefiniowanym w ontologii bazy wiedzy.
6. Transformacja wynikowej ontologii do modelu GIS służyć będzie przeniesieniu indywidualów opisanych w ontologii wynikowej do warstw geometrycznych GIS (obiekty przestrzenne + opis), z całym wzbogaconym zestawem atrybutów zawierających opis klasyfikacji wynikowej oraz nowe atrybuty wraz z wartościami, opisującymi wnioskowane cechy obiektów.
7. Wizualizacja wyników analiz, będących warstwami GIS, w postaci map tematycznych, wykorzystywać będzie tradycyjne narzędzia GIS (mogą to być narzędzia desktop GIS, jak i aplikacje sieciowe).

Architektura

Rozważana architektura systemu analiz i wspomaganie decyzji inspirowana jest opracowaniem OGC *Geospatial Semantic Web Interoperability Experiment Report* (Lieberman, 2006). Jednak przedstawione w raporcie GSW.IE rozwiązanie nastawione jest głównie na wyszukiwanie i przetwarzanie zasobów SDI. Rozwija ono metody eksploracji zasobów SDI, wzbogacając je o semantyczną interpretację danych, dzięki czemu możliwe jest uzyskiwanie informacji do ściśle określonych celów, spełniających specyficzne, wskazane przez użytkownika cechy, pozwalających następnie je przetwarzać i uzyskać odpowiedzi na złożone pytania.

Koncepcja, prezentowana w niniejszym artykule wykracza poza dotychczasowe korzystanie z infrastruktury SDI. Zmierza ona do wsparcia procesów decyzyjnych w zagadnieniach planowania przestrzennego i gospodarki przestrzennej, odnoszących się do realnych zjawisk, którymi te dziedziny zarządzają. Projektowane ontologie dziedzinowe modelują konkretną problematykę, a procesy decyzyjne są wspierane przez bazę wiedzy, zawierającą kryteria rozstrzygnięć. Tym co dodatkowo wyróżnia architekturę wspomaganie decyzji, to wykorzystanie filozofii DescriptionLogic do modelowania złożonych warunków i ograniczeń, zdefiniowanych w przepisach prawnych.



Rys. 1. Architektura systemu łączącego GIS i technologie sieci semantycznych

Przedstawiona na rysunku 1 architektura proponowanego rozwiązania składa się z następujących komponentów:

- 1) aplikacja Desktop GIS do edycji danych przestrzennych,
- 2) system zarządzania relacyjną bazą danych z rozszerzeniem geoprzestrzennym (np. PostgreSQL/PostGIS),
- 3) edytor ontologii zgodny ze standardem OWL2 do tworzenia ontologii opisujących model logiczny analizowanych zagadnień (własne ontologie dziedzinowe, baza wiedzy),
- 4) narzędzie do mapowania schematu bazy danych na model ontologii dziedzinowych,
- 5) narzędzie do transformacji danych przestrzennych do reprezentacji semantycznej,
- 6) silniki wnioskowania,
- 7) narzędzia transformacji semantycznych danych wynikowych na reprezentację warstw GIS.

Aplikacja Desktop GIS. Aplikacja będzie używana do czynności niewykraczających poza tradycyjne sposoby użytkowania tego typu aplikacji. Jej funkcjonalność zostanie wykorzystana do następujących zadań:

- pozyskania danych z serwisów WFS,
- zapewnienia dostępu do zasobów oraz funkcjonalności PostGIS,
- tworzenie i edycji własnych danych przestrzennych i gromadzenie ich w PostGIS.

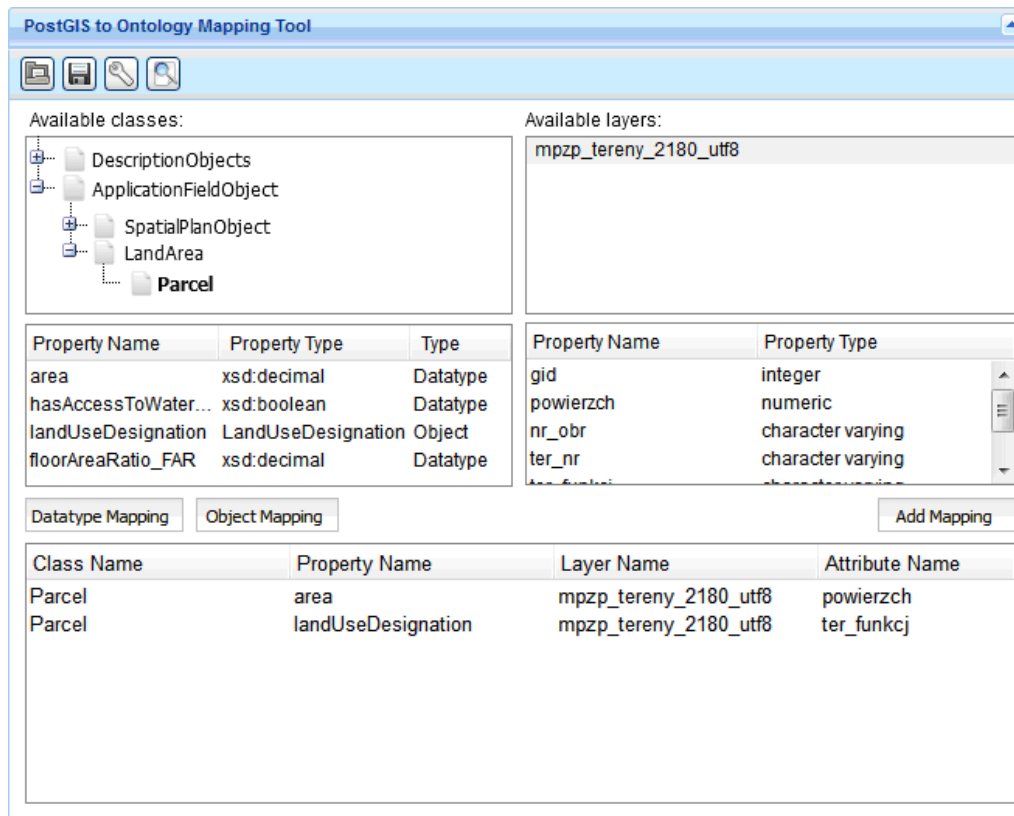
System zarządzania relacyjną bazą danych (RDMBS) z rozszerzeniem geoprzestrzennym. Może to być baza PostgreSQL/PostGIS, zaawansowany system bazodanowy wraz modułem geoprzestrzennym PostGIS, dostarczającym zgodny ze specyfikacją OGC zestaw geometrycznych typów danych oraz funkcji manipulowania geometrią i analiz przestrzennych. System bazodanowy z rozszerzeniem geoprzestrzennym będzie wykorzystany do przechowywania danych geometrycznych i opisowych, do prowadzenia analiz przestrzennych oraz generowania opisu topologicznego zasobów w ontologiach.

Edytor ontologii w standardzie OWL2. Aplikacja powinna zapewnić funkcje edycyjne i wizualizacyjne, jak na przykład Protégé. Aplikacja będzie wykorzystana do tworzenia ontologii opisujących model logiczny analizowanych zagadnień (własne ontologie dziedzinowe, baza wiedzy).

Narzędzie do mapowania schematu przestrzennej bazy danych na model ontologii dziedzinowych. Zasoby przestrzenne zawarte w bazie danych przestrzennych przechowywane są zgodnie z modelem relacyjnym według pewnej specyfikacji – schematu bazy danych. Określa on strukturę danych w formie tabel, ich ograniczenia oraz dozwolone operacje. Ze względu na dobrze poznane podstawy matematyczne oraz możliwość zadawania złożonych zapytań przeszukujących, model relacyjny jest szeroko wykorzystywany w bazach danych.

Warstwy obiektów przestrzennych przygotowanych w aplikacjach typu Desktop GIS zapisywane są najczęściej w postaci płaskich tabel zawierających kolumny geometryczne oraz pozageometryczne, które zwykle są typami prostymi. Każdy obiekt przestrzenny zawiera unikalny identyfikator tożsamy z kluczem głównym tabeli.

Aby umożliwić korzystanie z zasobów znajdujących się w relacyjnej bazie danych, w ramach systemów wnioskujących konieczne jest uzgodnienie schematu bazy z modelem ontologii dziedzinowych, a następnie przeniesienie zawartości bazy do postaci ABox. Przykładem istniejącego narzędzia umożliwiającego taką operację jest D2R, które pozwala na przeglądanie zawartości relacyjnych baz danych w postaci dokumentów RDF.



Rys. 2. Interfejs użytkownika narzędzia do mapowania schematu bazodanowego na model ontologii

W pierwszym etapie mapowanie polega na wskazaniu odpowiadających sobie tabel relacyjnej bazy danych oraz klas wczytanej ontologii. Następnie tworzony jest zestaw reguł mapowania odpowiednich kolumn tabeli na atrybuty klas, które mogą być podzielone na atrybuty obiektowe (ang. *ObjectProperty*) oraz atrybuty typów danych (ang. *DatatypeProperty*). W przypadku atrybutów typów danych zadanie jest trywialne i ogranicza się do sparowania typów danych relacyjnej bazy danych do typów zgodnych z XML-Schema. Atrybuty obiektowe wymagają stworzenia osobnych reguł przypisania obiektu do generowanego atrybutu indywiduum na podstawie wartości rekordu.

Przygotowanie reguł mapowania odbywa się poprzez wykorzystanie interfejsu użytkownika (rys. 2) lub ręczne stworzenie dokumentu XML.

Narzędzie transformacji danych przestrzennych do reprezentacji semantycznej pozwalają na przekształcenie danych gromadzonych w PostGIS do grafu RDF, tzw. ABox na podstawie przygotowanego wcześniej mapowania. Obiekty w wygenerowanych danych semantycznych są opisane zgodnie z problematyką modelowanej dziedziny gospodarki przestrzennej oraz zawierają właściwości opisujące relacje topologiczne pomiędzy obiektami zgodnie z ontologią NeoGeoSpatial. Dla realizacji tej funkcjonalności wykorzystuje się narzędzie SemGeo (Strzelecki, 2012). Po konwersji danych przestrzennych na dane semantyczne możliwe jest przeszukiwanie grafu poprzez wykonywanie zapytań SPARQL.

Interfejs protokołu SPARQL jest ustandaryzowanym sposobem dostępu do grafu RDF danych semantycznych wygenerowanych na podstawie danych przestrzennych i wyników analiz przestrzennych. Pozwala na przeszukiwanie grafu z wykorzystaniem języka zapytań SPARQL. Skonwertowane dane przestrzenne zostają zapisane do grafu RDF zgodnie z przyjętą ontologią. Silnik wykonywania zapytań SPARQL jest wzbogacony o możliwość wykorzystania funkcji przestrzennych zgodnych z NeoGeoSpatialOntology, dzięki czemu zasoby mogą być wyszukiwane poprzez testowanie zależności topologicznych występujących między obiektami przestrzennymi. Wygenerowane zasoby są publikowane przez udostępnienie punktu dostępowego SPARQL.

Silniki wnioskowania są wykorzystywane do reklasyfikacji obiektów na podstawie relacji przestrzennych i atrybutów. Obecnie najpowszechniej stosowane to Pellet i Fact++. Jest to istota całego procesu wspierania decyzji i odkrywania nowych faktów. Celem ich wykorzystania jest:

- 1) wnioskowanie o spełnieniu warunków decyzyjnych dla poszczególnych zasobów, obejmujących wiele warstw, powiązanych relacjami topologicznymi; definicja kryteriów jest zawarta w restrykcjach na właściwościach zdefiniowanych w tym celu klasach hierarchii modelu logicznego bazy wiedzy,
- 2) tworzenie ontologii wynikowych, jako TBox + ABox, zawierających obiekty w bogatszej klasyfikacji i z bogatszym zestawem właściwości.

Narzędzie transformacji danych z postaci ABox na reprezentację warstw GIS. Rezultaty wnioskowania zapisane są w ontologii wynikowej. Może ona być zwizualizowana jako graf RDF. Jednak dla analiz przestrzennych korzystniejsze jest przywrócenie jej modelu warstwy obiektów przestrzennych w stylu GIS/SDI. Taka warstwa może zostać przedstawiona w postaci mapy tematycznej lub kartodiagramu. Transformacja obejmuje:

- 1) uzupełnienie zapisu relacyjnego o dodatkowe atrybuty opisującą nową klasyfikację obiektów oraz o nowe właściwości wraz z ich wartościami,
- 2) umożliwienie wizualizacji w Desktop GIS faktów wynikających z wnioskowania w postaci mapy tematycznej lub kartodiagramu.

Baza wiedzy

Proces decyzyjny lub proces analizy, wykonywany przez projektowany system, realizowany jest na bazie zdefiniowanych w ontologiach modeli logicznych. Opisują one struktury taksonomiczne analizowanych zasobów, do których dokonywane są asercje obiektów i które są importowane do bazy wiedzy zawierającej kryteria decyzyjne. Ontologią jest również struktura wynikowa, zawierająca wywnioskowaną asercję indywiduów do klas, stanowiących rezultaty analiz wraz ze wzbogaconą listą właściwości.

W proponowanym rozwiązaniu informatycznym stosuje następujące grupy ontologii:

1. **Ontologia struktury danych.** Ontologia ta opisuje standardowy schemat zapisu obiektów reprezentujących zjawiska przestrzenne, ustalając sposób przedstawienia indywiduów w komponencie ABox. Stanowi ona wzorzec, do którego jest mapowany schemat bazodanowy: Tabela RDBMS → klasa obiektu, nazwa atrybutu/kolumna → właściwość, zawartość komórki w kolumnie → wartość właściwości.
2. **Ontologia zakresu tematyki.** Ontologie zestawiające wykaz dziedzin związanych z gospodarką przestrzenną, dla których są definiowane ontologie dziedziczne – nie są niezbędne, ale mają charakter informacyjny i porządkujący prezentowany obszar zainteresowań.

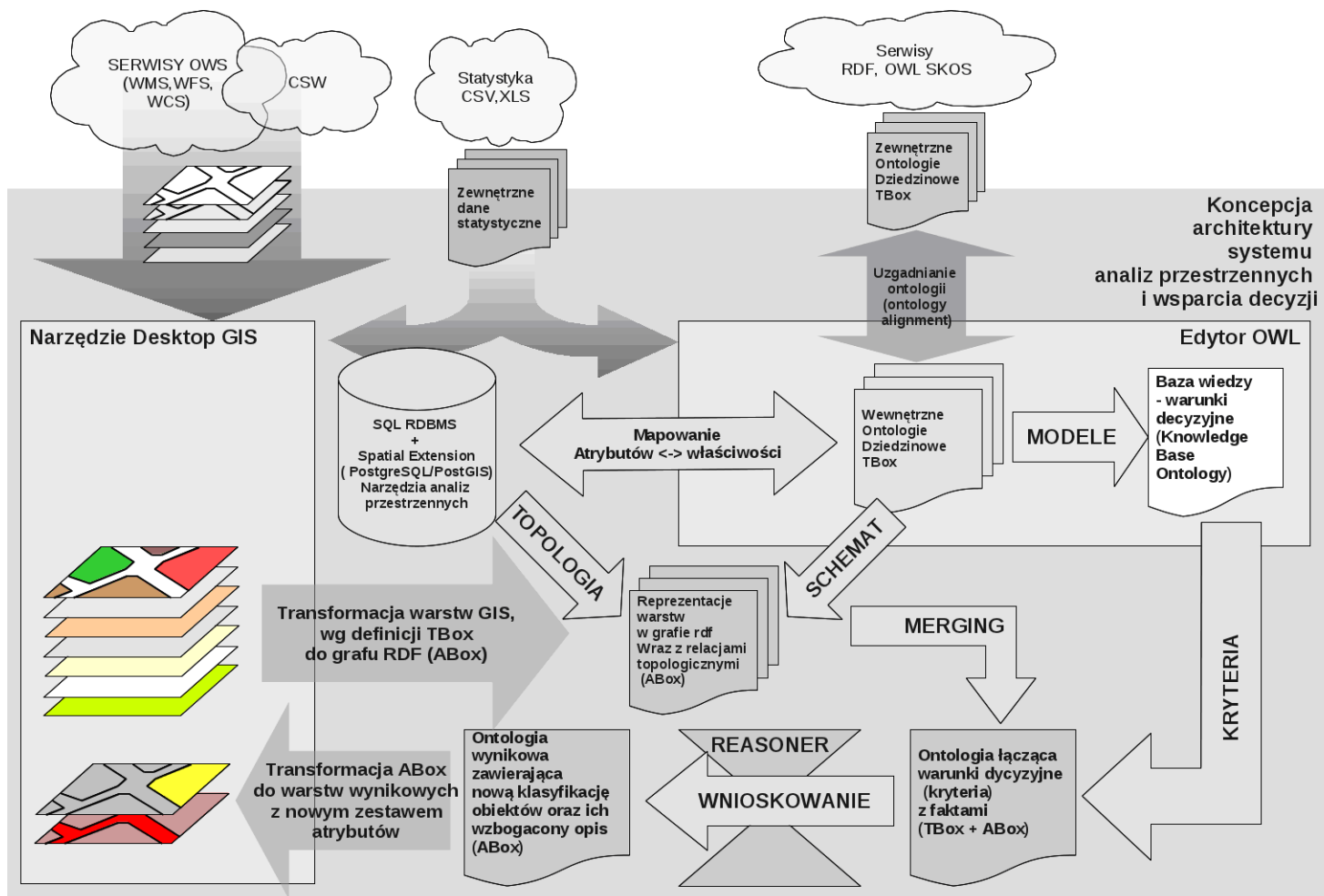
3. **Ontologie dziedzinowe.** Ontologie adresowane do poszczególnych dyscyplin gospodarki przestrzennej definiujące:
 - hierarchię obiektów – drzewo klas opisujące model pojęciowy danej dziedziny, który będzie rozszerzany w ontologii bazy wiedzy zawierającej kryteria decyzyjne,
 - hierarchię właściwości, z określeniem dziedziny (ang. *domain*) oraz zbioru wartości (ang. *range*),
 - zbiory wartości właściwości używanych do klasyfikacji obiektów z wykorzystaniem restrykcji.
4. **Ontologie decyzyjne.** Baza wiedzy odnosząca się do procesu decyzyjnego. Ontologie opisu reguł decyzyjnych do systemów wspierania decyzji, rozszerzające klasyfikację obiektów o klasy wynikowe ze zdefiniowanymi restrykcjami opisującymi warunki decyzyjne. Ontologie te nie będą natomiast wprowadzały nowych wartości dla właściwości.

Na rysunku 3 przedstawiono schemat przepływu danych dla implementacji procesu decyzyjnego zgodnie z zaproponowaną koncepcją. Zgodnie z tym schematem, w procesie wykorzystywane są zasoby danych przestrzennych i semantycznych z dwóch źródeł: zewnętrznych oraz tworzonych przez użytkownika (w procesie projektowym, w postępowaniu administracyjnym). Dane przestrzenne (typu GIS) są importowane z serwisów OGC/OWS lub tworzone w aplikacjach Desktop GIS oraz zachowywane w relacyjnej bazie danych posiadającej rozszerzenie geoprzestrzenne. Natomiast dane semantyczne są importowane z serwisów zewnętrznych lub tworzone w edytorze ontologii typu Protégé, a następnie integrowane przez techniki uzgadniania ontologii. Transformacja danych przestrzennych do reprezentacji semantycznej w postaci ABox, zawierającej indywidua obiektów przestrzennych wraz z relacjami topologicznymi wygenerowanymi w bazie danych z użyciem funkcji geoprzestrzennych, wymaga mapowania schematu bazodanowego na model semantyczny. Model zapisany w ontologiach dziedzinowych służy również jako szkielet budowanej bazy wiedzy, która dostarcza kryteriów decyzyjnych, łączonych później z zasobem ABox, wykorzystując technikę łączenia ontologii. Taka złączona ontologia stanowi informację wejściową dla silnika wnioskowania, który w procesie dokonuje reklasyfikacji obiektów i wzbogacenia opisujących ich właściwości. Otrzymana ontologia jest transformowana do wynikowych warstw GIS, które mogą być wykorzystane do zobrazowania rezultatów procesu decyzyjnego, w postaci map tematycznych i kartodiagramów w aplikacji Desktop GIS.

Podsumowanie

Możliwości zastosowania technologii semantycznych w obszarze GIS i SDI są szerokie. Zaproponowana przez autorów koncepcja stanowi przykład wykorzystania dotychczasowych osiągnięć i doświadczeń w tym obszarze, przekładających się na konkretne rozwiązania informatyczne. Koncepcję tę można traktować jako koncepcję systemu ekspertowego (w tym przypadku wspomagającego wydawanie decyzji o warunkach zabudowy), umożliwiającego korzystanie z wiedzy i ułatwiającego podejmowanie decyzji. Tworzone ontologie mogą być współdzielone i wykorzystywane w wielu niezależnych aplikacjach, co stanowi przewagę nad klasycznym systemem eksperckim, w którym wiedza gromadzona jest lokalnie.

Zaproponowane rozwiązanie stanowi krok w kierunku osiągnięcia interoperacyjności semantycznej w SDI, równocześnie zmierzając do budowy inteligentnej infrastruktury informacji przestrzennej.



Rys. 3. Proces wspomagania decyzji w projektowanym rozwiązaniu informatycznym

Literatura

- Andrei M., Berre A.J., Costa L., Duchesne P., Fitzner D., Grcar M., Hoffmann J., Klien E., Langlois J., Limyr A., Maue P., Schade S., Steinmetz N., Tertre F., Vasiliu L., Zaharia R., Zastavni N., 2008: SWING: An Integrated Environment for Geospatial Semantic Web Services. [In:] Bechhofer S., Hauswirth M., Hoffmann J., Koubarakis M. (eds.) 5th European Semantic Web Conference (ESWC 2008), LNCS 5021, Springer.
- Donato, Pasquale D., 2010: Geospatial Semantics: A Critical Review. By Part I, LNCS 6016 D. Taniar et al. (eds.) ICCSA 2010: 528–544.
- Fonseca F.T., Egenhofer M.J., Agouris P., Camara G., 2002: Using ontologies for integrated geographic information systems. *Transactions in GIS*: 231–257.
- Goczyła K., 2011, *Ontologie w systemach informatycznych*. Akademicka Oficyna Wydawnicza.
- Gruber R., 1993: A translation approach to portable ontology specification. *Knowledge Acquisition* No 5: 199-220.
- Iwaniak A., Kaczmarek I., Kubik T., Łukowicz J., Paluszyński W., Kourie D.G., Cooper A.K., Coetzee S., 2011: An Intelligent Geoportal for Spatial Planning. 25th International Cartographic Conference. Paris.
- Janowicz K., Schade S., Bröring A., Keßler C., Stasch C., Maué P., Diekhof T., 2009: A transparent semantic enablement layer for the geospatial Web. Terra Cognita 2009 Workshop in Conjunction with the 8th International Semantic Web Conference, <http://iswc2009.semanticweb.org/>, Washington, USA.
- Janowicz K., Keßler C., Bröring A., Stasch C., Schade S., 2010: Semantic enablement for Spatial Data Infrastructures. *Transactions in GIS*: 111-129.
- Jelokhani-Niaraki M., Malczewski J., 2012: A User-centered Multicriteria Spatial Decision Analysis Model for Participatory Decision Making: An Ontology-based Approach, GSDI 13 Quebec.
- Kaczmarek I., Łukowicz J., Cooper A., Iwaniak A., Kubik T., Paluszyński W., 2011: The use of ontologies in spatial planning domain. 25th International Cartographic Conference. Paris.
- Klien, E., Einspanier, U., Lutz, M., Hubner, S., 2004: An architecture for ontology-based discovery and retrieval of geographic information. *AGILE 2004*: 179-188.
- Kubik T., Iwaniak A., 2010: Building and maintaining metadata repositories with the aid of ontology tools and technologies. GSDI 12 World Conference. Singapore.
- Lieberman J., 2006: Geospatial Semantic Web Interoperability Experiment Report. Open Geospatial Consortium Inc.
- Lia W., Yanga C., Nebert D., Raskinc R., Houser P., Wua H., Lia Y., 2011: Semantic-based web service discovery and chaining for building an Arctic spatial data infrastructure, *Computers & Geosciences*: 1752-1762.
- Lutz, M., Klien, E., 2006: Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*.
- Łukowicz J., Kaczmarek I., Iwaniak A., 2012: Semantic metadata in SDI for decision support systems in spatial planning. Global Geospatial Conference 2012 „Spatially Enabling Government, Industry and Citizens”. Québec.
- Olfat H., Kalantari M., Rajabifard A., Williamson I., 2012: Towards a foundation for spatial metadata automation, *Journal of Spatial Science*: 65-81.
- Strzelecki M., 2012: Implementacja logiki biznesowej w systemach informacji przestrzennej z wykorzystaniem sieci semantycznych. Praca magisterska pod kierunkiem dr inż. Tomasza Kubika, Politechnika Wroclawska, Wrocław.
- Wache H., Voegelé T., Visser U., Stuckenschmidt H., Schuster G., Neumann H., Huber S., 2001: Ontology-based Integration of Information – A survey of existing approaches. *IJCAI*: 108-117.

Abstract

The paper presents the concept of an information system, which combines GIS with semantic web technologies. It relies on the integration of spatial data as well as information and knowledge about space, which defines the scope of the system. Thesauri and ontologies are used for this purpose. This approach enables to build intelligent systems for analysis and decision making in the field of spatial planning, including urban planning issues, the location of investment, property management, environmental protection and management of technical infrastructure and transportation. The second part of the paper describes development environment, which supports the implementation of systems developed according to the described concept. These solutions are an attempt of practical implementation of the idea of an intelligent spatial information infrastructure.

dr inż. Adam Iwaniak
adam.iwaniak@kon-dor.pl

mgr inż. Jaromar Łukowicz
jaromar.lukowicz@struktura.eu

mgr inż. Marek Strzelecki
marek.strzelecki@kon-dor.pl

mgr inż. Iwona Kaczmarek
iwona.kaczmarek@up.wroc.pl