

ONTOLOGIE W SDI ORAZ KONCEPCJA ICH WYKORZYSTANIA DO BUDOWY INTELIGENTNEGO GEOPORTALU*

ONTOLOGIES IN GIS AND THEIR USE FOR CONSTRUCTION OF INTELLIGENT GEOPORTALS

Adam Iwaniak¹, Tomasz Kubik²

¹ Instytut Geodezji i Geoinformatyki, Uniwersytet Przyrodniczy we Wrocławiu

² Instytut Informatyki, Automatyki i Robotyki, Politechnika Wroclawska

Słowa kluczowe: ontologia, informacja przestrzenna, interoperacyjność
Keywords: ontology, geospatial information, interoperability

Wprowadzenie

Pomimo bardzo dynamicznego rozwoju systemów GIS i budowy infrastruktur danych przestrzennych, powszechny dostęp do cyfrowych danych geoprzestrzennych jest dla przeciętnego użytkownika Internetu, pozbawionego wiedzy fachowej, mocno ograniczony. Utrudnienia wynikają z faktu, iż obecne portale geoinformacyjne koncentrują się na wyszukiwaniu i umożliwieniu dostępu do informacji przestrzennej, natomiast ich funkcjonalność w zakresie kontekstowego wyszukiwania, przetwarzania danych i kartograficznej prezentacji jest ciągle słabo rozwinięta.

Jedną z możliwości poprawy tej sytuacji jest rozwijanie technologii budowy inteligentnych geoportali, zorientowanych na wizualizację danych przestrzennych, wykorzystujących ontologie.

Ontologie mogą być wykorzystywane do realizacji różnych celów (Janowicz, Kessler, 2008; Kulik et al., 2005). Najczęściej stosowane są jako element inteligentnego interfejsu użytkownika umożliwiającego komunikację w języku naturalnym bądź zbliżonym do naturalnego. Użytkownik inteligentnego portalu formułuje zadanie, które jest następnie tłumaczone na postać języka ontologii sieci Web. Przejście w dziedzinę ontologii pozwala na dokonanie automatycznej analizy treści i zakresu zadania oraz wyprodukowanie wyniku. W procesie tym wykorzystywana jest wiedza zgromadzona w systemie, a w szczególności definicje pojęć dotyczących zagadnień geodezyjno-kartograficznych (Abdelmoty et al., 2005; Hess, de Vries, 2006; Lacasta et al., 2007).

Bazami wiedzy, które buduje się w oparciu o ontologie, są tezaury – dynamicznie, centralnie koordynowane słowniki. Dzięki przyjętej metodologii mogą one zawierać terminy i pojęcia ze wskazaniem relacji semantycznych występujących pomiędzy nimi. Na świecie

*Praca została częściowo zrealizowana w ramach projektu badawczego 4T12E01030.

tworzone są ogólnodostępne tezaury, które definiują taką strukturę informacji (Deliiska, 2007). Przykładem może być GEMET czy EUROVOC, tworzone przez Europejską Agencję Środowiska. Choć tezaury te nie są dostosowane do specyfiki geodezyjno-kartograficznej w Polsce, to jednak mogą być cennym zbiorem danych i wytycznych niezbędnych do stworzenia polskiego tezaurusa.

W niniejszym artykule opisano koncepcję inteligentnego geoportalu, w którym ontologie, tezaury i systemy eksperckie stanowią trzon kontekstowej usługi kartograficznej. Usługa ta ma stanowić uzupełnienie istniejącego wachlarza usług geoprzestrzennych OGC. Ważką kwestią poruszaną w artykule jest potrzeba zdefiniowania interfejsu takiej usługi oraz konieczność opracowania opisującego ją standardu.

Podstawy ontologii i sieci semantycznych

Paradygmaty rozwoju Internetu

Mówiąc o sieciach komputerowych, a w szczególności o Internecie, prowadzi się często rozważania nad ich modelem, rolą i kierunkiem rozwoju. Początkowo Internet funkcjonował jako system rozproszony, w którym zwykły użytkownik był biernym odbiorcą tego, co w sieci serwowały portale. To stadium Internetu nazwane zostało Web 1.0. W pewnym momencie nastąpiła rewolucyjna zmiana, polegająca na otwarciu Internetu na dane pochodzące od zwykłych użytkowników. W stadium tym, określanym jako Web 2.0, Internet stał się miejscem, gdzie zatarła się granica pomiędzy odbiorcą informacji a jej dostawcą.

Terminem Web 2.0 określa się serwisy internetowe powstałe po 2001 roku, w których generowanie treści przez użytkowników implementowane jest przy wykorzystaniu takich technologii jak XHTML, SOAP, AJAX, XML, RDF dających podstawy do tworzenia infrastruktury usług sieciowych web. W Web 2.0 zwykły użytkownik może zostać aktywnym współtwórcą treści zamieszczanych w sieci, z całym bagażem wynikających z tego faktu zalet i wad (patrz takie rozwiązania jak wikipedia, nasza klasa). Choć obecnie trwa jeszcze okres Web 2.0, już pojawiają się sygnały o nadchodzącej kolejnej rewolucji, określanej mianem Web 3.0.

Kolejna wersja rozwojowa Internetu, Web 3.0, bazować ma, podobnie jak Web 2.0, na technologiach związanych z XML i wykorzystujących usługi sieciowe. W implementacji Web 3.0 szczególną rolę odegrać ma sztuczna inteligencja, której wykorzystanie ma zrewolucjonizować sposób użytkowania informacji umieszczanych w sieci. Wyszukiwanie danych w Internecie polegać ma na czymś więcej niż na zwykłym ich indeksowaniu i użyciu słów kluczowych. Polegać ma ono na zrozumieniu informacji oraz lepszej, bardziej intuicyjnej ich prezentacji. Nowa jakość przetwarzania informacji osiągnięta ma być dzięki sieciom semantycznym, ontologiom, zastosowaniu języków naturalnych, metodom rozumienia i drażenia danych, uczeniu maszynowemu, technologiom agentowym i innym osiągnięciom. Dzięki zwiększonej szybkości połączeń sieciowych oraz inteligentnej analizie zapytań, znacząco poprawić się ma jakość procesu wyszukiwania. Możliwa ma być modularna budowa aplikacji sieciowych oraz wyraźnie zwiększony udział grafiki komputerowej w prezentowaniu informacji.

Web 3.0 to termin, który nie jest jeszcze do końca zdefiniowany. Używa się go jako określenia obrazującego przyszłą postać Internetu, skojarzone z nim różnego rodzaju działania oraz koncepcje prowadzące do konwersji obecnego systemu przekazu wiedzy do modelu

powszechnej bazy danych. W Internecie nowej generacji zawartość stron internetowych ma być reprezentowana przez pewien uniwersalny, rozszerzalny model pozwalający na ich analizę i przetwarzanie przez różne (w tym nieprzeładowane) aplikacje i systemy. Model ten korzysta z implementacji ontologii, w której obrębie znajdują się szeroko wykorzystywane metadane zapisane w języku RDF i jego aplikacjach, jak np. języku OWL.

Ontologia i jej języki

Istnieje szereg definicji ontologii. Według słownika Webstera ontologię definiuje się jako *dziedzinę metafizyki, która para się badaniem i wyjaśnianiem natury, jak i kluczowych właściwości oraz relacji rządzących wszelakimi bytami bądź głównych zasad i przyczyn bytu*. W ogólności ontologia zajmuje się odkrywaniem i opisywaniem „tego co jest”. Zajmuje się więc pewnym fragmentem rzeczywistości, mniej lub bardziej dokładnie określonym, próbując odpowiedzieć na pytania: „jak można wszystko sklasyfikować?”, „jakie klasy bytów są niezbędne do opisu i wnioskowania na temat zachodzących procesów?”, „jakie klasy bytu pozwalają wnioskować o prawdzie?”, „na podstawie jakich klas bytu można wnioskować o przyszłości?”

Ontologie są też pewnym formalizmem wykorzystującym kategoryzację oraz hierarchizację. Kategoryzacja to zdolność przyporządkowania symbolu do określonej grupy obiektów, posiadających określone cechy, np. „kot” należy do klasy kotów. Zestaw tych grup można określić jako zewnętrzny model pojmowania świata. Hierarchizacja natomiast jest umiejscowieniem określonej klasy w hierarchicznej strukturze.

W budowie modeli ontologii korzysta się często z podejścia obiektowego, tj. z możliwości definiowania całej struktury klas dziedziczących swoje właściwości oraz możliwością deklarowania obiektów tych klas (instancji klas). Obiektowość tę można zaobserwować w definicjach języków służących do zapisu ontologii. W większości tych języków fakty reprezentowane są podobnie: za pomocą klas, ich instancji i wzajemnych związków. Grupa formalnych języków zapisu ontologii nie jest zbyt liczna. Do grupy tej zaliczane są takie języki jak: Ontobroker, SHOE (*Simple HTML Ontology Extensions*), OIL (*Ontology Inference Layer lub Ontology Interface Layer*), DAML (*DARPA Agent Markup Language*) + OIL (*Ontology Inference Layer*), OWL (*Ontology Web Language*). W przeważającej większości języki te korzystają z RDF i RDFS (*Resource Description Framework Schema*). Dla starszych języków możliwa jest translacja do/z KIF (*Knowledge Interchange Format*), który bez problemu może być tłumaczony do/z RDF. Szczególne znaczenie dla implementacji usług sieciowych mają języki DAML z DAML-S oraz OWL z OWL-S służące do opisu usług sieciowych.

Obecnie istnieją na świecie ontologie zawierające bogate zestawy definicji klas i właściwości. Przykładem może być ontologia OpenCyc, która w wersji 1.0 zawierała już ponad 6000 zdefiniowanych klas i ponad 60 000 właściwości.

OIL

Ontologia OIL jest strukturą składającą się z kilku części, które również mogą posiadać własną strukturę, niektóre z nich są opcjonalne, inne powtarzalne. Dla opisu ontologii w OIL wyróżnia się poziomy:

- poziom obiektu, na którym opisywane są konkretne wystąpienia ontologii,

- pierwszy poziom meta, na którym umieszczane są rzeczywiste definicje ontologiczne. Tu definiuje się terminologię, która może być uprzedmiotowiona na poziomie obiektu. Jest to zasadniczy poziom OIL (nazywany także definicją ontologii), będący narzędziem służącym tworzeniu ustrukturalizowanych słowników o dobrze zdefiniowanej semantyce.
- drugi poziom meta (tzn. poziom meta-meta), nazywany też kontenerem ontologii, służy opisowi cech ontologii, takich jak: autor, nazwa, przedmiot itp. Dla wyrażania metadanych ontologii często używa się Dublin Core Metadata Element Set (<http://www.dublincore.org/>).

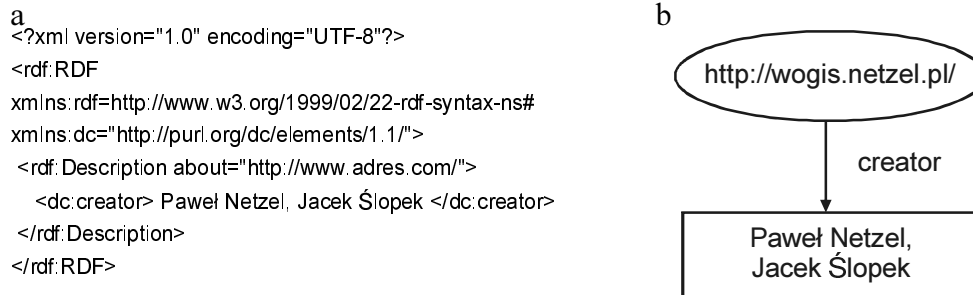
Opis danej ontologii umieszcza się w pliku tekstowym. Podstawowymi elementami takiego pliku są definicje klas (**class-def**) oraz cech (**slot-def**). Definicje klasy łączą nazwy klasy z ich opisami. W pliku ontologii mogą pojawić się takie opcjonalne elementy jak: **type** – definicja typu, który może być zarówno podstawowy, jak i definiowalny; podstawowe klasy pozwalają na tworzenie niezbędnych warunków dla przynależności do klasy, **subclass-of** – lista jednego lub kilku wyrażen klasy (**class-expression**). Klasa zdefiniowana w definicji klasy musi być podklasą każdego wyrażenia klasy na liście. **Slot-constraint** – lista zero lub więcej ograniczeń cech (**slot-constrain**). Definiowana klasa, dla której wypisane są ograniczenia musi być podklasą każdego ograniczenia na liście (zauważmy, że **slot-constraint** definiuje klasę). **Class-expression** może być nazwą klasy, ograniczeniem cech lub arbitralnie utworzonym zestawem połączeń boole'owskich **class-expressions**. **Slot-constraint** jest listą jednego lub więcej ograniczeń stosowanych do cechy. Cecha jest relacją binarną (tzn. jej wystąpieniami są pary jednostek), ale **slot-constraint** to faktyczne definicje klasy – jego wystąpieniami są te jednostki, które spełniają ograniczenia.

RDF

Język RDF (ang. *Resource Description Framework*) jest aplikacją XML służącą do reprezentacji danych na potrzeby programów katalogujących. Celem opracowania tego języka było: zwiększenie zakresu możliwości przeglądarek internetowych i wyszukiwarek w obszarze automatycznego pozyskiwania informacji; ułatwienie katalogowania zawartości i relacji między danymi w danej witrynie WWW; umożliwienie „inteligentnemu” oprogramowaniu udostępnienie i wymianę informacji.

RDF rozszerza standardowy model informacyjny XML o konstrukcję wyższego rzędu, umożliwiającą tworzenie takich stwierdzeń, jak np.: „Tomasz powiedział, że adresem poczty elektronicznej dla zasobu &o2 jest jacek@gmail.com”. Podstawą tych konstrukcji są trójczłonowe wyrażenia opisujące, odpowiednio: podmiot – predykat – wartość (czasem nazywane też: podmiot – orzeczenie – dopełnienie). Konstrukcje te reprezentowane są w składni RDF przez element główny, przestrzenie nazw, element **Description** z atrybutem **about** (mogą pojawić się też formy skrócone) oraz pojemniki, pozwalające łączyć elementy w pewne grupy.

Opis zasobów w języku RDF to nic innego jak zestaw metadanych (rys. 1). Podczas opracowywania metadanych występuje problem określenia dziedziny używanych deskryptorów. Podobnie jak w przypadku tworzenia metadanych dla danych przestrzennych, aby urzeczywistnić realizację wdrożenia ontologii tworzone są profile metadanych – tj. zbiory zawierające definicje wszystkich używanych w danej dziedzinie deskryptorów. W przypad-



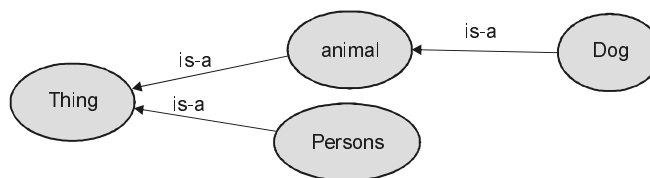
Rys. 1. Przykład dokumentu RDF oraz wizualizacja jego zawartości o następującej interpretacji: twórcą (creator) zasobu <http://wogis.netzel.pl/> jest Paweł Netzel, Jacek Ślopek

ku danych przestrzennych model metadanych bazuje na deskryptorach zdefiniowanych w różnych normach, z czego najbardziej popularną jest ISO 19115 zawierająca 300 deskryptorów. Jednak ani 300 ani 3000 deskryptorów nie jest w stanie opisać wszystkich możliwych obiektów, które mogą się pojawić na stronach WWW.

W RDF, podobnie jak każdej innej aplikacji XML, można definiować przestrzenie nazw, w których umieszczane mogą być nowe deskryptory. Każdy plik RDF, podobnie jak każdy inny plik XML, może skorzystać z tych definicji przez zamieszczenie w swoim elemencie głównym deklaracji odpowiednich przestrzeni nazw. Rozwój prac nad metadanymi i RDF wykazał, iż RDF nie jest wystarczający do reprezentacji ontologii, co istotnie ogranicza możliwości wyszukiwania i wnioskowania. Dalsze prace doprowadziły do opracowania języka OWL i OWL-S pozwalającego na opis usług sieciowych.

OWL

Język OWL jest językiem wyższego poziomu bazującym na XML oraz RDF. Język ten wywodzi się z języka DAML+OIL. Istnieją trzy rodzaje języka OWL: OWL Lite, OWL DL oraz OWL Full. Podobnie jak język OIL, język OWL umożliwia definiowanie klas i właściwości. W stosunku do RDF posiada on istotne rozszerzenie polegające na „dołożeniu logiki pierwszego rzędu” (*description logic*). Pozwala to wykorzystywać ontologie nie tylko jako zestaw metadanych, ale również jako bazę wiedzy, na której przeprowadzać można proces wnioskowania. W konsekwencji ontologie wyrażone w OWL są w stanie udzielić odpowiedzi na dwa zasadnicze pytania: „Co to jest x? Czy to co mam to x?”. Przykładowe definicje wyrażone w języku OWL zebrano w tabeli 1. Definicje te pochodzą z dokumentu draft200.owl zawierającego definicję ontologii opisującej dziedzinę SOA (tj. zagadnień związanych z architekturą zorientowaną na usługi). Dokument ten można znaleźć na stronie internetowej organizacji OpenGroup pod adresem: <http://www.opengroup.org/projects/soa-ontology/>.



Rys. 2

Zasadniczo siła wyrazu języków zapisu ontologii jest podobna. Można to zaobserwować porównując zapis tej samej ontologii w diagramie klas jak na rysunku 2 w języku OWL i OIL (tab. 2).

Tabela 1. Przykładowe definicje klasy, podklasy i właściwości wyrażone w języku OWL (fragment ontologii dla SOA zaproponowanej przez OpenGroup)

Opis	Przykład
Definicja klasy Actor	<code><owl:Class rdf:ID="Actor"/>.</code>
Definicja podklasy HumanActor dziedziczącej po klasie Actor	<code><owl:Class rdf:ID="HumanActor"> <rdfs:subClassOf rdf:resource="#Actor"/> </owl:Class></code>
Definicja właściwości isDescribedBy	<code><owl:ObjectProperty rdf:ID="isDescribedBy"> <rdfs:range rdf:resource="#Description"/> <owl:inverseOf rdf:resource="#describes"/></owl:ObjectProperty></code>

Tabela 2. Porównanie zapisu ontologii zawierającej definicje trzech klas w różnych językach

Format OWL	Format OIL
<pre><?xml version="1.0" encoding="ISO-8859-1"?> <rdf:RDF xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#"> <owl:Ontology rdf:about=""/> <owl:Class rdf:about="#Persons"/> <owl:Class rdf:about="#Dog"> <rdfs:subClassOf> <owl:Class rdf:about="#animal"/> </rdfs:subClassOf></owl:Class></rdf:RDF></pre>	<pre>begin-ontology ontology-container title "x" creator "x" description "x" description.release "1.01" type ontology format "pseudo-xml" identifier "file:/" language "OIL" language "en-uk" ontology-definitions class-def Persons class-def Dog subclass-of animal end-ontology</pre>

Różnice pomiędzy słownikiem, taksonomią, tezauresem, ontologią i meta-modelem

Słownik (a dokładniej kontrolowany słownik) jest uporządkowaną listą terminów kontrolowaną przez instytucję upoważnioną do rejestracji nowych terminów i modyfikację terminów istniejących (wg <http://www.metamodel.com/article.php?story=20030115211223271>).

Wszystkie terminy w kontrolowanym słowniku powinny mieć jednoznaczne, związane definicje. Rejestracja powinna bazować na następujących założeniach: jeśli jakiś termin używany jest zwykle do wyrażania różnych koncepcji w różnych kontekstach, to problem jego niejednoznaczności rozwiązuje się jawnie, deklarując wszystkie złożoności tego terminu; jeśli wiele terminów jest stosowanych do identyfikowania tej samej rzeczy, jeden z tych terminów powinien być terminem referencyjnym, zaś pozostałe powinny być zdefiniowane jako synonimy albo aliasy.

Taksonomia jest kolekcją terminów z kontrolowanego słownika zorganizowanych w strukturę hierarchiczną. Każdy termin w taksonomii występuje w jednej lub więcej relacjach typu

rodzic-dziecko względem innych terminów tej taksonomii. Relacje te mogą mieć różny charakter lub typ (np. całość-część, klasa-instancja), jednak dobrą praktyką jest ograniczenie tych relacji do jednokrotnego dziedziczenia. Kontrolowany słownik może zawierać terminy bez opisu znaczenia (np. dla ogólnie przyjętych terminów) lub z bardzo szczegółowym opisem. Taksonomia dostarcza dodatkowych znaczeń dla terminów poprzez ich hierarchizację. W tradycyjnej taksonomii znaczeniem jest generalizacja/specjalizacja lub związek typu „jest częścią”.

Tezaurus jest kolekcją terminów z kontrolowanego słownika powiązanych w sieć. Sieć ta tworzona jest przez związki dziedziczenia (tj. relacje rodzic-dziecko), które mogą być jednokrotne i wielokrotne (choć zazwyczaj dziedziczenie jest jednokrotne) oraz różne asocjacje (całość-część, klasa-instancja). Dzięki sieciowemu charakterowi tezaury możliwe jest znajdowanie w nim terminów „powiązane ze sobą”, a nie tylko terminów pozostających w tym samym obszarze taksonomii. Siła asocjacji terminów w tezaurysie może być niejednorodna, tzn. może być tak prosta, jak związek typu „A odpowiada B”, lub bardziej złożona, jak związek typu „A jest podobne do B, ale różne od C”. Liczba zdefiniowanych związków i asocjacji w tezaurysie jest ograniczona.

Taksonomie i tezaury mogą łączyć terminy w kontrolowanym słowniku poprzez dziedziczenie i asocjacje, jednak nie posiadają reguł gramatycznych, które określałyby ograniczenia na użycie tychże terminów do wyrażania bądź modelowania czegoś znaczącego w obrębie dziedziny zainteresowań. Jeśli taksonomia posiada duży wybór bardzo starannie zdefiniowanych znaczeń dla hierarchicznych związków, wtedy jest ona mocno podobna do ontologii.

Termin ontologia bywa używany w różnych znaczeniach: jako glosariusz i słownik danych, tezaurus i taksonomia, schemat i model danych, jako formalizm pozwalający na typowanie i wnioskowanie. Formalna ontologia jest kontrolowanym słownikiem wyrażonym w języku ontologii z możliwością dynamicznej rozbudowy. Język ontologii posiada gramatykę pozwalającą na wykorzystanie terminów słownika do wyrażania pojęć w pełnym kontekście określonej dziedziny. Gramatyka zawiera formalne ograniczenia (dotyczące struktury wyrażań, asercji, zapytań), zasadę łącznego używania terminów z kontrolowanego słownika ontologii, sposób wprowadzania nowych terminów i definiowania nowych zależności. Restrykcje wynikające z gramatyki ontologii mogą być rygorystyczne bądź luźne. Częstość gramatyki „lekkich” ontologii nie jest do końca zdefiniowana, np. może posiadać niejawnie reguły, które nie są jawnie udokumentowane. Ontologie otwarte są na rozbudowę – teoretycznie mogą zawierać nieskończoną liczbę zdefiniowanych klas, ich właściwości i związków.

Meta-model jest jawnym modelem pewnych konstrukcji i reguł potrzebnych do budowy specyficznych modeli w obrębie dziedziny zastosowań. Poprawny meta-model jest ontologią, ale nie wszystkie ontologie są zamodelowane jawnie jako meta-modele. Meta-model może być przeglądany z trzech różnych perspektyw: 1) jako zbiór elementów konstrukcyjnych i reguł używanych do budowy modeli, 2) jako model specyficznej dziedziny zainteresowań, 3) jako instancja innego modelu.

Jeśli stworzono ontologię, która jest zbiorem klas i relacji, zaś do stworzenia zbioru danych (instancji klas oraz asercji opisujących wzajemne relacje zgodnie ze słownikiem) i zbiór ten interpretowany ma być jako model danej dziedziny, to ontologia jest meta-modelem, zaś stworzony zbiór jest modelem.

Wspólne cechy słowników, taksonomii, tezaurysów, ontologii i meta-modele można opisać w następujący sposób: – wszystkie są pewnym podejściem wspomagającym strukturyzację, klasyfikację, modelowanie i reprezentację koncepcji i relacji odnoszących się do jakie-

gość tematu znaczącego dla pewnej społeczności; – dedykowane są do umożliwienia społeczności osiągnięcia konsensusu w sprawie używania tych samych terminów w ten sam sposób; – istnieje zbiór terminów, które zaakceptowane są przez społeczność jako baza referencji do tych koncepcji i związków; – znaczenie terminów jest wyrażane w pewien określony sposób do pewnego stopnia szczegółowości; – istnieją rozmyte, czasem niejasne notacje używane na wiele różnych sposobów przez różnych użytkowników i społeczności.

Podstawowe różnice pomiędzy wyżej wymienionymi abstraktami to: – siła dostarczonego opisu znaczenia terminów; – notacja lub język, który został użyty do objaśniania znaczenia; – sposób użycia wykorzystywanych elementów (jedna rzecz może być inaczej wykorzystywana w różnych taksonomiach, tezaurusach, ontologiach, meta-modelu); – możliwość rozbudowy.

Aplikacje bazujące na ontologiach

Tezaurusy

Implementacje tezaurusów występują, jako zbiory semantycznie i hierarchicznie powiązanych terminów, ułatwiające wyszukiwanie pochodnych informacji lub słowniki wyrazów bliskoznacznych dołączane do niektórych procesorów tekstu. Tezaurusy zazwyczaj są specjalizowane do obsługi pojęć z wybranej dziedziny (Deliiska, 2007). Na rysunku 3 przedstawiono przykład tezaurusu GEMET (*General Multilingual Environmental Thesaurus*) opracowanego na zlecenie Europejskiej Agencji Środowiska, zawierającego ponad 6000 pojęć opracowanych w 22 językach.

Semantyczne wyszukiwarki

Najpopularniejsze wyszukiwarki internetowe, tj. aplikacje przeznaczone do automatycznego wyszukiwania informacji i danych w Internecie, bazują głównie na mechanizmie indeksowania i wykorzystania słów kluczowych. Taka metodologia pozwala zaspokoić potrzeby zwykłych użytkowników. W przypadku wyszukiwania informacji geoprzestrzennych sytuacja ma się inaczej. Ze względu na naturę danych przestrzennych należy użyć dedykowanego dla tych danych mechanizmu. Obecnie w Internecie można napotkać dane przestrzenne, które coraz częściej opisywane są przez metadane wyrażane w języku XML zgodnie z profilami metadanych bazującymi na normie ISO 19139. Pojawiają się też serwery katalogowe, w których metadane te są gromadzone i wyszukiwane dzięki zaimplementowaniu mechanizmów wyszukiwania pożądanej informacji przestrzennej.

Wyszukiwanie w sieciach Web 3.0 polegać ma na automatycznej analizie zapytań oraz automatycznym przeglądaniu ontologii bądź metadanych zapisanych w formacie RDF bezpośrednio na stronach internetowych, jak również w bazach danych. W założeniach wyszukiwanie powinno opierać się o inteligentną analizę zapytań, jak również generowanie odpowiedzi znaczeniowo związanych z zapytaniem (związanych semantycznie, a nie tylko poprzez słowa kluczowe). Stąd przeglądarki Web 3.0 nazywane są semantycznymi wyszukiwarkami. Ich praktyczne możliwości można prześledzić na przykładzie wyszukiwarek Swogle i hakia (rys. 4).

Ontologie w SDI

Prace nad ontologiami w budowie SDI trwają od kilku lat. Ontologie stosowane są głównie do reprezentacji wiedzy organizacji oraz zwiększenia automatyzacji procesu użycia różnych zbiorów danych w semantycznie spójny sposób. Innym celem, dla którego tworzone są ontologie jest usystematyzowanie pojęć używanych przy tworzeniu modeli i standardów. Przykładem ontologii opisujących zagadnienia związane z topografią są ontologie opracowane przez brytyjski Ordnance Survey (<http://www.ordnancesurvey.co.uk/oswebsite/ontology/>).

W głównym nurcie zastosowań ontologii znalazły się również próby semantycznego opisanie geoinformacyjnych usług sieciowych w celu automatyzacji procesu ich wyszukiwania zgodnie z oczekiwaniami użytkowników poprzez użycie języka zapytań zbliżonego do języka naturalnego. Podobne akcje podejmowane są dla danych przestrzennych (Batcheller, 2008; Lutz, 2007; Schuurman, Leszczyński, 2006). Nie chodzi tu o zmianę formy komunikacji na język naturalny, ale zdolność wnioskowania i wykorzystania wyrażen bliskoznacznych. Opisanie usługi za pomocą metadanych zgodnych z normą ISO 19119 lub języka WSDL pozwala co prawda wyszukać usługę po słowach kluczowych oraz precyzyjnie opisać syntaktykę jej interfejsu. Jednak ten sposób opisu niesie z sobą zbyt mało informacji aby dokonać automatycznego wyszukania i uruchomienia usługi spełniającej kryteria użytkownika. Złożoność problemu można dostrzec w prozaicznym zadaniu wyszukania usługi pozwalającej obliczyć pole wieloboku przy zadanych współrzędnych jego wierzchołków. W zależności od tego, czy pole liczone ma być jako pole powierzchni na sferze lub czy pole powierzchni rzutu punktów na płaszczyznę należałoby wywołać inną usługę. Niestety, klasyczny opis usług zawiera tylko definicje interfejsów usług (metody z listą parametrów), na podstawie których nie da się wywnioskować, jak faktycznie wykonane będą obliczenia. Można powiedzieć, że klasyczny opis usługi dostarcza syntaktykę wywołań metod, a nie ich semantykę. Zadaniem dostarczenia semantyki może natomiast wykonać zaimplementowana ontologia. Dzięki jej użyciu można dokonywać nawet automatycznej konstrukcji całego łańcucha usług (Lemmens et al., 2007; Yue et al., 2007).

Trudną rolę brokera geoportalu wykorzystującego klasyczny opis usług zilustrować można kolejnym przykładem. Klasyczne przeszukiwanie serwerów katalogowych odbywa się po słowach kluczowych opisanych w normach, np. ISO 19115 i ISO 19119, przy czym istotną rolę odgrywają nie tylko nazwy atrybutów ale ich wartości, które konsekwentnie dla wszystkich wystąpień powinny być jednakowe. Z powyższego powodu zapytania serwera katalogowego typu „podaj mi wszystkie **mapy zasadnicze** dla powiatu wrocławskiego **wykonane** przez OPGK” oraz „podaj mi wszystkie **podstawowe mapy kraju** dla powiatu wrocławskiego **opracowane** przez OPGK” są różne i z dużym prawdopodobieństwem dadzą różne wyniki, w przeciwieństwie do wyszukiwarek semantycznych.

Wykorzystanie ontologii daje znacznie szersze możliwości bowiem pozwala na reprezentację wiedzy i umożliwia wykonanie na jej podstawie wnioskowania. Załóżmy, iż lotnisko, na którym może wylądować Jumbo Jet, musi mieć co najmniej cztery kilometry długości i 50 metrów szerokości. Pytanie „podaj mi miasta w Polsce do których można dolecieć samolotem typu Jumbo Jet” w klasycznej wyszukiwarce musiałoby zostać zamienione na pytanie do bazy danych o odpowiedniej składni, np.: „znajdź wszystkie X gdzie X zawiera Y i Y.PasStartowy.DługoscPasaStartowego>4000 m i Y.PasStartowy.SzerokoscPasaStartowego>50 m”. Dla wyszukiwarek semantycznych pierwotne zapytanie mogłoby pozostać niezmienione – to silnik ontologii automatycznie rozpoznałby istotę zapytania (i przekazałby je dalej).

Oprócz automatycznej obsługi języka naturalnego (lub zbliżonego do naturalnego) ontologie zapewniają jeszcze dodatkową możliwość. Otóż w odróżnieniu od klasycznego podejścia do zadawania pytań i wyszukiwania odpowiedzi ontologiczne podejście zwalnia użytkownika z obowiązku posiadania szczegółowej wiedzy o dziedzinie zapytania. To wyszukiwarka i ontologie taką wiedzę muszą posiadać i dostarczać ją użytkownikowi.

Inteligentne geoportale

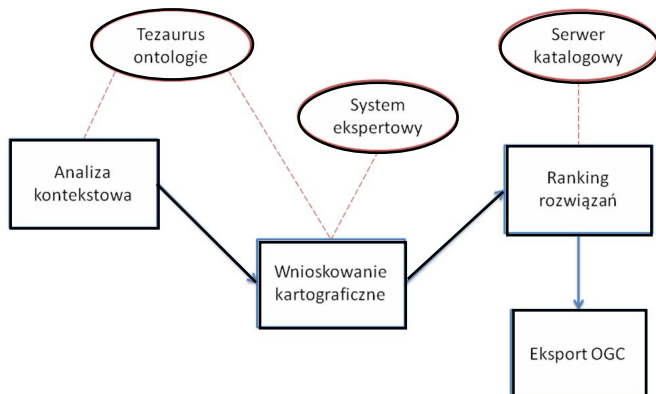
Opracowanie ontologii dla wybranej dziedziny może skutecznie zwiększyć zdolność komunikacji człowieka z komputerem oraz „podnieść poziom inteligencji” budowanych systemów. Poniżej autorzy przedstawiają skrótowy opis fragmentu projektu systemu – „inteligentnego geoportalu”, który mógłby się przyczynić do znacznego zwiększenia wykorzystania danych pochodzących z państwowego zasobu geodezyjnego i kartograficznego.

Jednym z zadań inteligentnego geoportalu ma być wspomaganie procesu wyszukania danych przestrzennych w sieci Internet oraz doboru poprawnych (akceptowalnych) metod wizualizacji danych ze względu na cel opracowania. Portal adresowany ma być do mało zaawansowanych użytkowników Internetu, posiadających ograniczoną wiedzę o istniejących danych przestrzennych – w szczególności o danych pochodzących z zasobu geodezyjnego i kartograficznego i ich jakości.

Wprowadzenie ontologii do SDI wymagałoby wprowadzenia nowych komponentów do architektury geoportalu (rys. 5). W architekturze tej powinny pojawić się przynajmniej cztery dodatkowe moduły:

- Analiza kontekstowa,
- Wnioskowanie kartograficzne,
- Ranking rozwiązań,
- Eksport OGC.

Moduł *Analiza kontekstowa* odpowiadać ma za właściwe zrozumienie żądań użytkowników przez system. W tym celu wykorzystuje się tezaurus i ontologie opracowane dla zasobu



Rys. 5. Schemat funkcjonalny fragmentu „inteligentnego portalu” odpowiedzialnego za opracowanie kartograficznej wizualizacji danych przestrzennych

danych przestrzennych. Na podstawie rozpoznanego celu, tematu, zakresu i szczegółowości żadanego opracowania moduł *Wnioskowanie kartograficzne* ma za zadanie dobrać właściwą metodę kartograficzną, legendę oraz styl. Realizacja tego zadania wspierana ma być, oprócz wbudowanych mechanizmów wnioskowania bazujących na ontologiach, implementacją dedykowanego systemu ekspertowego. Umożliwi to wykorzystanie heurystyk, które są stosowane do rozwiązywania problemów o

podobnej charakterystyce i złożoności. Zbiór możliwych rozwiązań zawiera zazwyczaj kilka wariantów, dlatego moduł *Ranking rozwiązań* musi dokonać oceny i wytypować najlepsze z nich, uwzględniając przy tym dostępność danych. Wymusza to współpracę z serwerem katalogowym. Ostatnim krokiem jest wygenerowanie stylów prezentacji poszczególnych warstw tematycznych zgodnie ze specyfikacjami OGC, w tym przypadku WMC (*Web Map Context*) i SDL (*Styled Layer Descriptor*). Realizacją tego zadania zająć ma się moduł *Export OGC*. Export do formatów opracowanych przez OGC zapewni interoperacyjność systemu, tj. możliwość jego współpracy z geoportalem klasy INSPIRE.

Zakończenie

Metadane i ontologie znajdują się w głównym nurcie rozwoju Internetu i są podstawą sieci Web w wersji 3.0. Pracują nad nimi wszystkie największe korporacje: Google, Microsoft, Yahoo. Ich głównym zadaniem jest porządkowanie zawartości Internetu i uczynienie go bardziej zrozumiałym dla systemów informatycznych, co stanowi podstawę wyszukiwania kontekstowego. Warto jednak zwrócić uwagę, że opracowanie metadanych, a tym bardziej ontologii, jest trudne i kosztowne. Można więc wnioskować, że ich opracowanie nawet dla niewielkiego procentu zasobu sieci Internet należy rozpatrywać w wieloletniej perspektywie.

Ontologie stały się kluczowym elementem w semantycznym opisie usług sieciowych, co stanowi podstawę do ich automatycznego wyszukiwania i integracji. Należy mieć nadzieję, że wielkość zainteresowania i wysokość nakładów na badania w tym kierunku zaowocują rozwiązaniami pozwalającymi na ich łatwą implementację na szeroką skalę.

W opisaney koncepcji inteligentnego geoportalu autorzy zaproponowali wykorzystanie ontologii do budowy interfejsu użytkownika z możliwością definiowania zapytań w języku zbliżonym do naturalnego. Dzięki temu komunikacja z portalem jest szybsza, bardziej intuicyjna i zwalnia użytkownika z posiadania szczegółowej wiedzy na temat danych przestrzennych, serwerów katalogowych i kartografii. Takie rozwiązanie powinno przyczynić się do spopularyzowania i znacznie szerszego wykorzystania istniejących zasobów danych przestrzennych, a w tym danych z Państwowego Zasobu Geodezyjnego i Kartograficznego.

Literatura

- Abdelmoty A. I., Smart P. D., Jones C. B., Fu G., Finch, D., 2005: A critical evaluation of ontology languages for geographic information retrieval on the Internet, *Journal of Visual Languages and Computing, Elsevier Science*, 16, 331-358.
- Batcheller J. K., 2008: Automating geospatial metadata generation -An integrated data management and documentation approach, *Computers and Geosciences, Elsevier Science*, 34, 387-398.
- Deliiska B., 2007: Thesaurus and Domain Ontology of Geoinformatics, *Transactions in GIS*, 11, 637-651.
- Hess C., de Vries M., 2006: From models to data: A prototype Query Translator for the cadastral domain Computers, *Environment and Urban Systems, Elsevier Science*, 2006, 30, 529-542.
- Janowicz K., Kessler C., 2008: The Role of Ontology in Improving Gazetteer Interaction International, *Journal of Geographical Information Science (IJGIS)*, 10.
- Kulik L., Duckham M., Egenhofer M., 2005: Ontology-driven map generalization, *Journal of Visual Languages and Computing, Elsevier Science*, 16, 245-267.
- Lacasta J., Nogueras-Iso J., Béjar R., Muro-Medrano P., Zarazaga-Soria F., 2007: A Web Ontology Service to facilitate interoperability within a Spatial Data Infrastructure: Applicability to discovery, *Data & Knowledge Engineering, Elsevier Science*, 63, 947-971.

- Lemmens R., de By R., Gould M., Wytzisk A., Granell C., van Oosterom P., 2007: Enhancing Geo-Service Chaining through Deep Service Descriptions, *Transactions in GIS*, 11, 849-871.
- Lutz, M., 2007: Ontology-Based Descriptions for Semantic Discovery and Composition of Geoprocessing Services GeoInformatica, Kluwer Academic Publishers-Plenum Publishers, 2007, 11, 1-36.
- Schuurman N., Leszczynski A., 2006: Ontology-Based Metadata, *Transactions in GIS*, 10, 709-726.
- Yue P., Di L., Yang W., Yu G., Zhao P., 2007: Semantics-based automatic composition of geospatial Web service chains, *Computers and Geosciences, Elsevier Science*, 33, 649-665.

Abstract

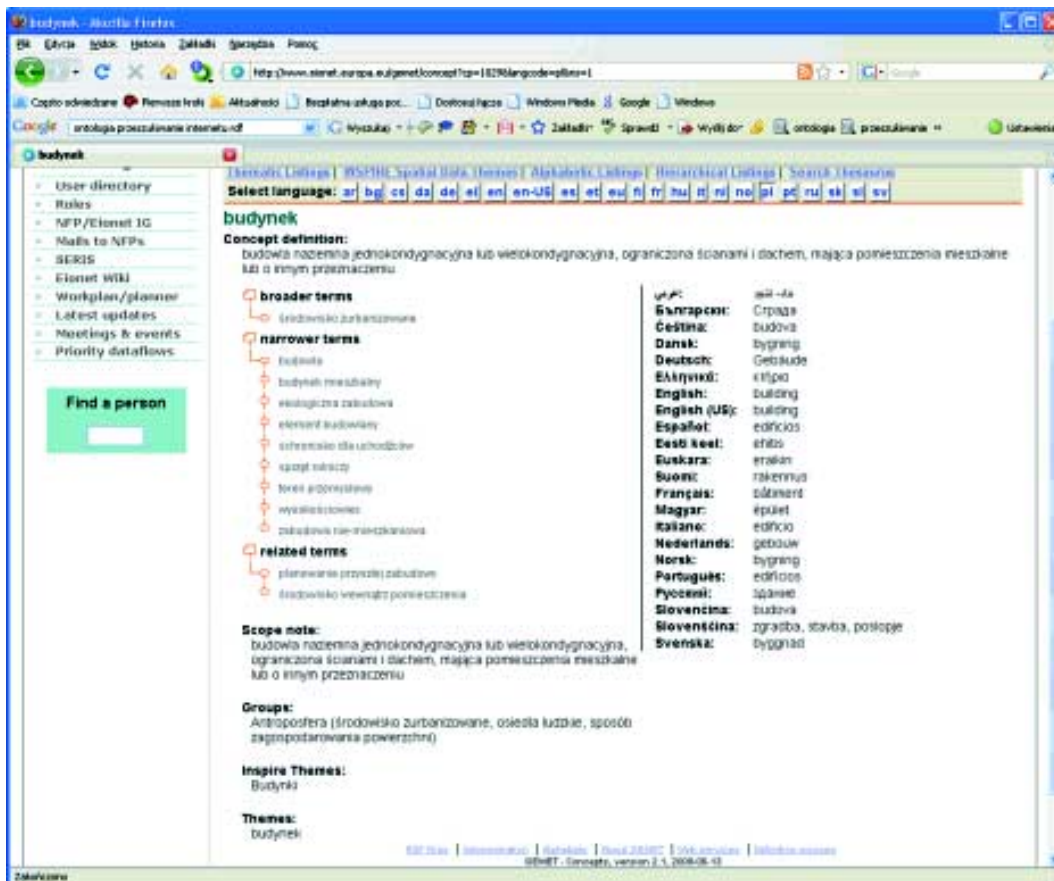
Looking at the development of Internet we find that the role of metadata and ontology has been increasing day by day. Next generation of websites and internet utilities will offer to the user more intelligent interfaces, able to interpret queries and provide results more adequate to his or her needs. The development of the semantic web (or Web 3.0) has come into focus of main software providers such as Google, Microsoft, Yahoo. These corporations try to organize Internet accessible resources in such a way that they will be easier for searching, interpreting and understanding by information systems equipped with context search engines. The potential costs of ontology implementation and the difficulties which may arise cannot be neglected. It may turn out that ontology implementation in real life will take quite a long time, even when applied to a small percentage of Internet resources.

The ontology may be used in Web 3.0 in several ways. As the main component of the semantic web services ontology may support implementation of the user interface with natural language, it may be used for automatic searching and integration service, etc. We may predict that the interest in ontology development and its use in Internet applications will grow significantly.

The concept of an intelligent geoportal described in the paper uses ontology and thesauri for semantic querying spatial data. In this approach, ontology enriches the description of data sources or services, providing them with machine-interpretable semantic context. Thanks to ontology-based design, logical reasoning can be used for interpreting user queries and for supporting results ranking. Thus, the concept provides the bases for the solution where user-geoportal interactions can be faster, more intuitive, and easier to manage even for users who do not have any specific knowledge of geospatial data, catalogue services and cartography. The solution proposed should encourage broader use of existing spatial data and other geodetic and cartographic service resources.

dr inż. Adam Iwaniak
adam.iwaniak@up.wroc.pl

dr inż. Tomasz Kubik
tomasz.kubik@pwr.wroc.pl



Rys. 3. Strona internetowa z tezaurem GEMET

a



b



Rys. 4. Przykłady wyszukiwarek bazujących na ontologiach: a – hokia, b – Swoogle