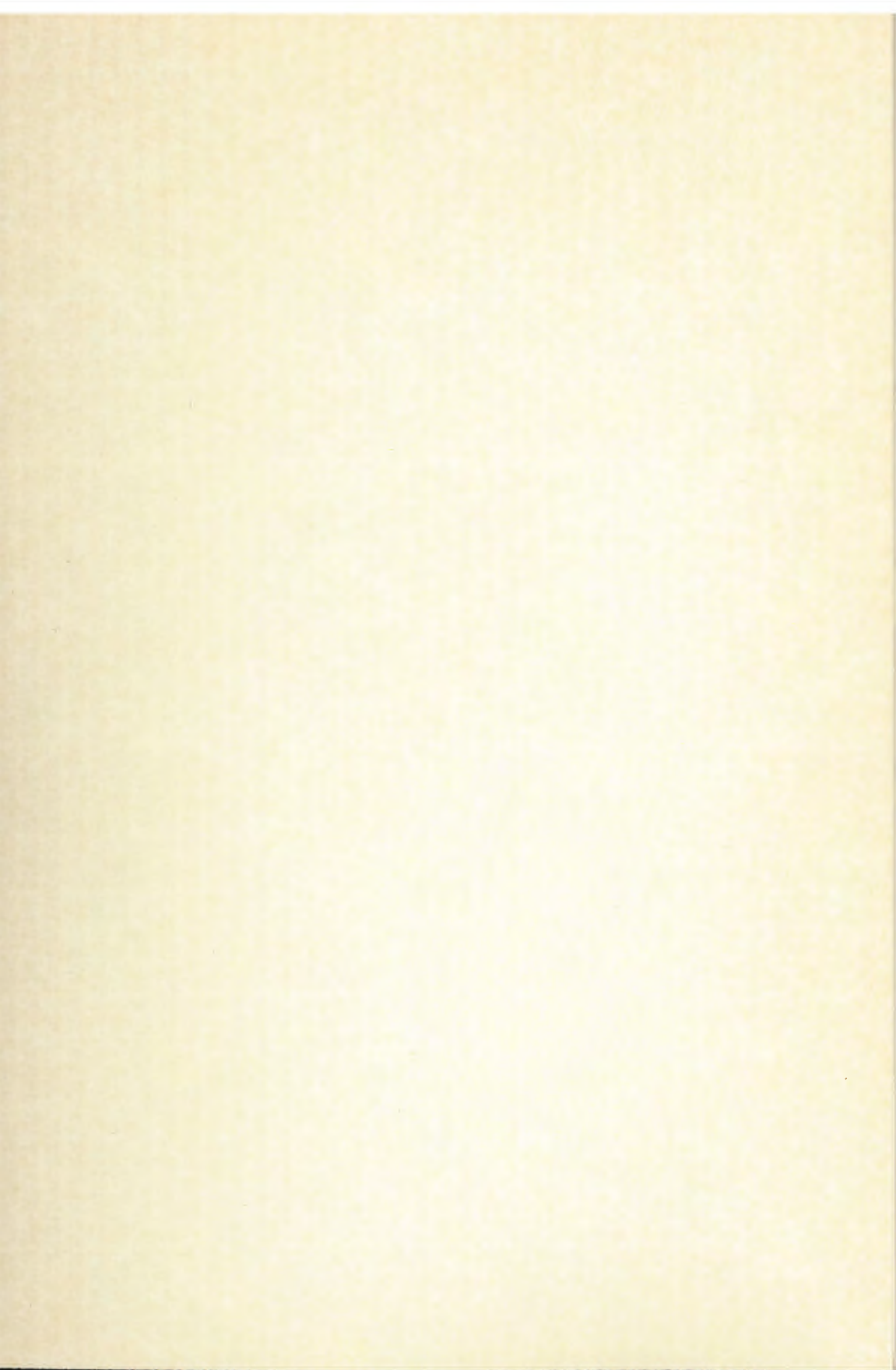




POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

**WSPOMAGANIE INFORMATYCZNE
ROZWOJU
SPOŁECZNO-GOSPODARCZEGO
I OCHRONY ŚRODOWISKA**

Redakcja:
Jan Studziński
Ludostław Drelichowski
Olgierd Hryniewicz





**WSPOMAGANIE INFORMATYCZNE
ROZWOJU
SPOŁECZNO-GOSPODARCZEGO
I OCHRONY ŚRODOWISKA**

Polska Akademia Nauk Instytut Badań Systemowych

Seria: BADANIA SYSTEMOWE

tom 36

Redaktor naukowy:

Prof. dr hab. Jakub Gutenbaum

Warszawa 2004

**WSPOMAGANIE INFORMATYCZNE
ROZWOJU
SPOŁECZNO-GOSPODARCZEGO
I OCHRONY ŚRODOWISKA**

Redakcja:

Jan Studziński
Ludosław Drelichowski
Olgierd Hryniewicz

Książka wydana dzięki dotacji KOMITETU BADAŃ NAUKOWYCH

Książka zawiera wybór artykułów poświęconych omówieniu aktualnego stanu badań w kraju w zakresie rozwoju modeli, technik i systemów zarządzania oraz ich zastosowań w różnych dziedzinach gospodarki narodowej. Wyodrębnioną grupę stanowią artykuły omawiające aplikacyjne wyniki projektów badawczych i celowych KBN.

Recenzenci artykułów:

Dr Lucyna Bogdan
Prof. dr hab. inż. Olgierd Hryniewicz
Dr Grażyna Petriczek
Prof. dr hab. inż. Andrzej Straszak
Dr inż. Jan Studziński



Senia 45187

Komputerowa edycja tekstu: Anna Gostyńska

© Instytut Badań Systemowych PAN, Warszawa 2004

Wydawca: Instytut Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa

Sekcja Informacji Naukowej i Wydawnictw IBS PAN
tel. 836-68-22

Druk: Zakład Poligraficzny Urzędu Statystycznego w Bydgoszczy
Nakład 110 egz.

ISBN 83-85847-92-8
ISSN 0208-8028

SYSTEM DYSTRYBUCJI DOKUMENTÓW UNIJNYCH W POLSCE Z JAWNĄ REPREZENTACJĄ WIEDZY W POSTACI ONTOLOGII

Klaudia A. Ławcewicz

Studium Doktoranckie przy Instytucie Badań Systemowych PAN

Sławomir Zadrozny

Wyższa Szkoła Informatyki Stosowanej i Zarządzania

<Sławomir.Zadrozny@wsisz.edu.pl>

An ontology based system supporting document distribution is proposed. The heart of the system is an ontology that explicitly represents knowledge on recipients' interests as well as domain knowledge. The system is meant to solve a real-life problem of the distribution of UE documents among the Polish government institutions. The ontology, written in OWL, was prepared and applied for document categorization using Protégé and RACER tools.

Keywords: Ontology, knowledge representation, UE documents distribution, U32Mail, OWL, Protégé, RACER.

1. Wstęp

Przedmiotem niniejszej pracy jest eksperymentalny schemat wspomagania dystrybucji dokumentów UE, przesyłanych w ramach systemu U32Mail (Ławcewicz, 2003), do odpowiednich urzędów (resortów) w oparciu o wiedzę jawnie reprezentowaną w postaci ontologii. Wiedza ta dotyczy zarówno preferencji poszczególnych urzędów, co do tematyki otrzymywanych dokumentów, jak i powiązań pomiędzy pojęciami odnoszącymi się do poszczególnych kategorii tematycznych.

W związku z małym doświadczeniem Polski odnośnie uczestnictwa w procesie decyzyjnym UE, nie wykształcono w kraju precyzyjnego podziału kompetencji polskich urzędów zgodnego z klasyfikacją tematyczną stosowaną w UE. Stworzenie mechanizmu klasyfikacji dokumentów umożliwiającego automatyczną dystrybucję wiadomości do adresatów na terenie Polski jest więc niezwykle pożądane. W niniejszej pracy przedstawiono schemat dystrybucji dokumentów oparty na jawnej reprezentacji wiedzy w postaci ontologii. Automatyczna klasyfikacja dokumentów w proponowanym podejściu zasadza się na wnioskowaniu na podstawie ontologii z zastosowaniem reguł logik deskryptywnych.

W dalszej części omawiamy krótko pojęcie ontologii, a następnie przedstawiamy narzędzia wykorzystane do tworzenia ontologii i do prowadzenia wnioskowania w oparciu o nią. W kolejnych punktach, szczegółowo opisana została

skonstruowana ontologia U32MailOnt i eksperyment obliczeniowy dotyczący automatycznej dystrybucji dokumentów testowych.

2. Pojęcie ontologii

Ontologię można określić jako system formalnej specyfikacji terminów wykorzystywanych w określonym obszarze tematycznym i relacji zachodzących pomiędzy nimi. Tak rozumiana ontologia znajduje zastosowanie w zadaniach takich jak: reprezentacja i zarządzanie wiedzą, przetwarzanie języka naturalnego czy wyszukiwanie informacji tekstowej (ang. *information retrieval*).

Podstawowe cele, którym służyć ma tworzenie ontologii to (Noy, McGuinness, 2001):

- współdzielenie przez ludzi i agentów programowych uzgodnionej interpretacji terminologii dotyczącej danej dziedziny;
- możliwość wielokrotnego użycia wiedzy z danej dziedziny;
- jawne określenie założeń dotyczących danej dziedziny jakie przyjęto przy konstruowanie danego systemu informatycznego;
- oddzielenie wiedzy dziedzinowej od wiedzy operacyjnej (ogólniejszej);
- analiza wiedzy dziedzinowej.

Pojęcie ontologii nawiązuje do innych, wcześniej stosowanych metod reprezentacji wiedzy takich jak *sieci semantyczne* czy *ramy*. Obecnie jednak, jako podstawę teoretyczną ontologii przyjmuje się powszechnie *logiki deskryptywne*. Logiki deskryptywne można określić jako rodzaj uproszczonego - w stosunku do klasycznej logiki predykatów - formalizmu reprezentacji wiedzy, umożliwiającego tworzenie modelu pojęciowego (konceptualizacji) określonej dziedziny przez zdefiniowanie istotnych *pojęć* (klas), powiązań między nimi oraz wykorzystania ich do opisu obiektów występujących w danej dziedzinie. Wspomniane uproszczenie ogranicza możliwości opisowe logik deskryptywnych, ale jednocześnie zapewnia ich rozstrzygalność. Pojęcia podstawowe reprezentowane są klasycznie jako predykaty. Dodatkowo, wprowadzone są *konstruktory* pozwalające na tworzenie definicji pojęć złożonych. Wśród nich występują oczywiście klasyczne spójniki logiczne jak również inne specyficzne dla logik deskryptywnych konstruktory. Z poszczególnymi pojęciami mogą być związane pewne ich *właściwości*, będące formalnie relacjami określonymi na iloczynie kartezjańskim przestrzeni dwóch pojęć (np. właściwość „bycia matką” jest relacją określoną na iloczynie kartezjańskim przestrzeni pojęć KOBIETA i OSOBA). Podstawową zależnością pomiędzy pojęciami jaką reprezentuje się w ramach ontologii jest *subsumpcja*, czyli relacja określająca zawieranie się w sobie przestrzeni dwóch pojęć, „bycia podklasą” (np. przestrzeń pojęcia MATKA zawiera się w przestrzeni pojęcia OSOBA). Traktując ontologię jako bazę wiedzy zwykle wyróżnia się jej dwie części: TBox

opisujący zależności pomiędzy pojęciami i ABox opisujący przede wszystkim *instancje* poszczególnych pojęć.

Ontologie, odwołując się do logik deskryptywnych, przejmują dostępny w nich efektywny mechanizm wnioskowania. Przykładowo, możliwe jest wnioskowanie o przynależności konkretnych obiektów do przestrzeni poszczególnych pojęć (czyli ustalanie, którego pojęcia instancją jest dany obiekt).

Dla zapisu ontologii stworzono specjalizowane języki. Wśród nich, najbardziej popularnym wydaje się obecnie OWL, którym posługujemy się również w tej pracy.

3. Narzędzia wykorzystane do konstrukcji i eksploatacji ontologii

Dwa podstawowe narzędzia, którymi posłużyliśmy się dla realizacji postawionego zadania to Protégé i RACER. Protégé (*Protégé User's Guide*), opracowany w Stanford University, stanowi zintegrowane środowisko wspomagające tworzenie systemów z jawną reprezentacją wiedzy w postaci ontologii. Protégé, wprost lub poprzez dodatkowe moduły, wspiera wiele języków służących do zapisu ontologii; wśród nich DAML+OIL i OWL. Umożliwia również wykonywanie wielu operacji na tworzonej, testowanej, bądź istniejącej ontologii. Szeroka funkcjonalność Protégé wynika z możliwości dołączania do bazowej aplikacji różnorodnych nakładek. Na potrzeby niniejszej pracy wykorzystano następujące nakładki: Protégé-OWL¹, umożliwiającą budowanie ontologii w oparciu o język OWL, oraz ezOWL², pozwalającą na graficzną prezentację systemów ontologicznych zaprojektowanych przy pomocy języka OWL.

System RACER³ jest „maszyną wnioskującą” dla baz wiedzy opartych na logice deskryptywnej. RACER wspiera większą część funkcjonalności KRSS⁴. Oferuje typowe usługi wnioskowania w odniesieniu zarówno do TBox jak i ABox. W szczególności, RACER umożliwia (Haarslev, Moller) uzyskanie odpowiedzi na następujące pytania:

- w odniesieniu do komponentu TBox ontologii:
 - czy zbiór obiektów opisanych przez dane pojęcie nie jest pusty (czyli ustalanie niesprzeczności pojęć względem TBox),
 - czy zbiór obiektów opisanych przez dane pojęcie jest podzbiorem zbioru obiektów opisanych przez inne pojęcie (czyli ustalanie subsumpcji pojęć; szerzej: możliwe jest określenie całej hierarchii pojęć),

¹ <http://protege.stanford.edu/plugins/owl/>

² <http://iweb.etri.re.kr/ezowl/>

³ Reasoner for ABoxes and Concept Expressions Renamed

⁴ Knowledge Representation System Specification, www.krss.org

- w odniesieniu do komponentu ABox:
 - do której, najbardziej specyficznej, klasy zdefiniowanej w TBox należy dany obiekt (czyli klasyfikacja obiektów).

Popularnym graficznym interfejsem dla RACER-a jest RICE⁵. Z jego pomocą użytkownik formułuje komendę i przesyła ją do realizacji do RACER-a. Wyniki komend są ilustrowane w postaci drzewa pojęć zdefiniowanych w ontologii, listy instancji poszczególnych klas itp.

W systemie Protégé usługi wnioskowania są domyślnie skonfigurowane do pracy z serwerem RACER. Pozwalają one na sprawdzenie niesprzeczności wybranego pojęcia, to jest ustalenie czy istnieją instancje tego pojęcia. Możliwa jest również weryfikacja hierarchii pojęć: system automatycznie wykrywa zależność subsumpcji pomiędzy poszczególnymi pojęciami.

4. Koncepcja automatycznej dystrybucji dokumentów

Każdy dokument przychodzący do Polski z UE jest opisany zbiorem kodów określonych kategorii tematycznych. Wszystkie urzędy, które miały zostać przyłączone w pierwszym etapie do polskiej gałęzi systemu U32Mail określiły, którymi kodami są zainteresowane, wraz z rozróżnieniem „kodów wiodących” (to jest takich, których wystąpienie w dokumencie powinno zdecydować o przesłaniu takiego dokumentu do danej instytucji jako wiodącej w procesie jego przetwarzania) oraz „kodów współpracujących” (to jest takich, których wystąpienie w dokumencie powinno zdecydować o przesłaniu takiego dokumentu do danej instytucji jedynie jako współpracującej w procesie jego przetwarzania). Na tej podstawie skonstruowano ontologię U32MailOnt, reprezentującą wiedzę na temat preferencji zaangażowanych instytucji.

Ponadto, aby podnieść efektywność automatycznej klasyfikacji dokumentów wprowadzono do ontologii dodatkową warstwę umożliwiającą reprezentację semantyki kodów tematycznych. Warstwę tę stanowią pewne ogólne kategorie (terminy kluczowe), które korespondują ze znaczeniem poszczególnych kodów tematycznych. Dzięki temu, zarówno kody tematyczne, kategorie jak i powiązania pomiędzy nimi występujące są jawnie reprezentowane w ramach ontologii. W przypadku zmian w warstwie znaczeniowej korekcje podlega jedynie baza wiedzy (ontologia), a cały system dystrybucji funkcjonuje bez zmian.

Ontologia zaproponowana w niniejszej pracy do wsparcia automatycznej dystrybucji dokumentów została skonstruowana następująco. Dla każdego urzędu definiuje się klasę dokumentów, które wchodzą w krąg jego zainteresowania. Dokładniej, definiuje się dla każdego urzędu kilka takich klas odpowiadających dokumentom dla których urząd jest:

⁵ RACER Interactive Client Environment

- jednostką wiodącą,
- jednostką wspierającą,
- kandydatem na jednostkę wspierającą.

Poszczególne klasy zdefiniowane są z użyciem konstruktorów klas, tak aby odpowiadało to preferencjom wyrażonym wcześniej przez urzędy. Wszystkie te klasy są podklasami ogólnej klasy DOKUMENT. Klasa ta ma właściwość maKOD (o wartościach z klasy KOD, obejmującej kody tematyczne używane w systemie), za pomocą której podaje się kody tematyczne wymienione w danym dokumencie. Z kolei, poszczególne kody opisane są za pomocą terminów kluczowych właściwych dla dziedziny do której odnosi się dany kod. Terminy powiązane są znaczeniowo co znajduje odzwierciedlenie w ontologii i pozwala na pełniejszą reprezentację zależności semantycznych pomiędzy kodami.

Następnie, dla dokumentu który ma zostać zaklasyfikowany tworzona jest instancja klasy DOKUMENT o właściwościach maKOD określonych stosownie do kodów tematycznych faktycznie występujących w klasyfikowanym dokumencie (w ontologii, dowolna właściwość może przyjmować dla dowolnej instancji wiele wartości, gdyż właściwość jest faktycznie relacją określoną na iloczynie kartezyjańskim stosownych klas, tu: klas DOKUMENT i KOD). Po utworzeniu instancji uruchamiane jest wnioskowanie w ontologii, które skutkuje przypisaniem nowoutworzonej instancji do stosownych klas. Na podstawie tego przypisania decyduje się o przesłaniu dokumentu do konkretnych urzędów.

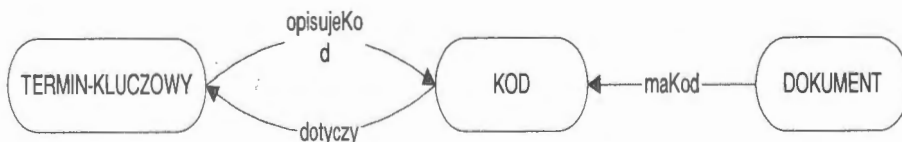
5. Ontologia U32MailOnt

System ontologiczny zaproponowany do rozwiązania problemu obejmuje trzy podstawowe klasy:

- DOKUMENT
- KOD
- TERMIN-KLUCZOWY

i cztery właściwości:

- maKod (\subseteq DOKUMENT \times KOD)
- dotyczy (\subseteq KOD \times TERMIN-KLUCZOWY)
- opisujeKod (\subseteq TERMIN-KLUCZOWY \times KOD)
- pokrewny (\subseteq TERMIN-KLUCZOWY \times TERMIN-KLUCZOWY)



Rysunek 1. Podstawowe klasy i właściwości ontologii U32MailOnt

W ramach powyższych klas stworzone zostały podklasy oraz dodane instancje reprezentujące wszystkie dane zebrane w resortach (opisujące ich preferencje) oraz pewną wiedzę dotyczącą dziedzin opisywanych kodami tematycznymi.

Klasa KOD reprezentuje wszystkie kody tematyczne używane w systemie U32MAIL. Jej podklasy odnoszą się do wszystkich resortów (urzędów) biorących udział w procesie dystrybucji i dzielą się na dwie grupy:

- Kody4[nazwa resortu]-Wiod (reprezentują „kody wiodące” dla danego resortu),
- Kody4[nazwa resortu]-Wsp, (reprezentują „kody współpracujące” dla danego resortu).

Klasa KOD, i co za tym idzie, wszystkie jej podklasy posiadają właściwość „dotyczy”, która odnosi się do klasy TERMINY-KLUCZOWE. Właściwość ta łączy instancje kodów tematycznych z odpowiednimi terminami kluczowymi.

Każda z podklas klasy KOD jest zdefiniowana, w terminach OWL-a, z pomocą opisu klasy typu wyliczeniowego, co w Protege zapisuje się, np. dla klasy Kody4MS-Wsp, jako:

{COHOM ∪ COJUR ∪ CONSOM ∪ COTER ∪ COUR ∪ CRIMORG ∪ CRS-CRP ∪ JUR ∪ JURINFO ∪ LEX ∪ OJ-CRP1 ∪ OJ-CRP2 ∪ OJ-RGEM ∪ PI ∪ PTS-A ∪ PV-CONS}.

a zapisane w OWL z użyciem składni RDF/XML wygląda następująco

```

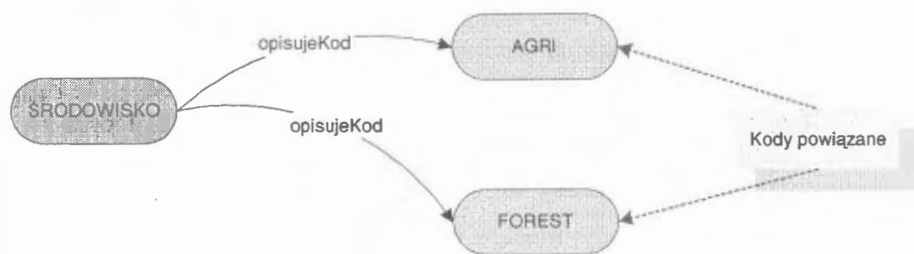
DEFINICJA KLASY Kody4MS-Wsp:
<owl:Class rdf:ID="Kody4MS-Wsp">
  <owl:equivalentClass>
    <owl:Class>
      <owl:oneOf rdf:parseType="Collection">
        <KOD rdf:about="#COHOM"/>
        <KOD rdf:about="#COJUR"/>
        ....
        ....
        ....
        <KOD rdf:about="#PTS-A"/>
        <KOD rdf:about="#PV-CONS"/>
      </owl:oneOf>
    </owl:Class>
  </owl:equivalentClass>
  <rdfs:subClassOf rdf:resource="#KOD"/>
</owl:Class>
    
```

Celem do wprowadzenia do ontologii pojęcia (klasy), TERMIN-KLUCZOWY jest pełniejsza reprezentacji semantyki poszczególnych kodów tematycznych. Pozwala to na uniknięcie sytuacji, w której zgodnie z preferencjami wprost wyrażonymi przez urzędy, do sprawy związanej z dokumentem przypisać można zbyt mało lub zgoła żaden resort. Zakładamy wtedy, że kody opisane tymi samymi terminami

kluczowymi (jednym lub większą ich liczbą) powinny być znaczeniowo bliskie sobie. Jeśli dany kod tematyczny nie został wskazany przez żaden resort, to szukamy kodów znaczeniowo bliskich i wybieramy resorty nimi zainteresowane.

Z klasą TERMIN-KLUCZOWY związane są dwie właściwości: „opisujeKod” i „pokrewny”. Właściwość „opisujeKod” jest właściwością odwrotną do właściwości „dotyczy” zdefiniowanej dla klasy KOD i przyjmuje jako wartości kody związane z danym terminem kluczowym. Właściwość „pokrewny” wiąże ze sobą poszczególne instancje terminów kluczowych, które są znaczeniowo podobne (np., ekonomia i gospodarka). Właściwość „pokrewny” jest właściwością symetryczną.

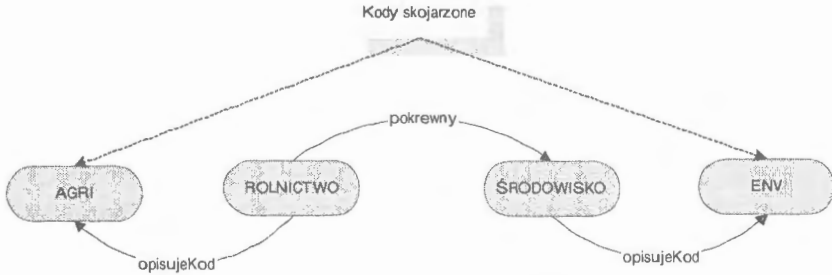
Dzięki powyższym właściwościom możliwe jest kojarzenie dokumentów z resortami na dwa dodatkowe sposoby (oprócz powiązania *explicite* przez przecięcie zbioru kodów współpracujących czy wiodących ze zbiorem kodów przypisanych dokumentowi) – dokumenty mogą być *skojarzone* bądź *powiązane* z konkretnym resortem. Dokument jest powiązany z danym resortem, jeśli kod tematyczny występujący w dokumencie jest opisany tym samym terminem kluczowym co choć jeden kod współpracujący tego resortu. Jako przykład weźmy dokument opatrzony kodem FOREST. Nie występuje on wśród kodów wspierających resortu MRiRW (Ministerstwo Rolnictwa i Rozwoju Wsi). Jednak kody AGRI i FOREST opisane są tym samym terminem kluczowym „ŚRODOWISKO”, w związku z czym taki dokument będzie dokumentem „powiązanym” z resortem MRiRW (nazwiemy też powiązanymi kody FOREST i AGRI).



Rysunek 2. Kody powiązane

Dokument jest skojarzony z danym resortem, jeżeli kod występujący w dokumencie jest opisany terminem kluczowym, który jest pokrewny do terminu opisującego co najmniej jeden kod związany *explicite* (czyli wskazany przez resort jako „kod współpracujący”) z tym resortem. Analogicznie jak poprzednio, również takie kody nazwiemy skojarzonymi.

Powiązania tego typu są słabsze niż *explicite* wskazane przez zainteresowane instytucje, i w związku z tym powinny być wykorzystywane tylko w wypadkach, gdy wnioskowanie zgodnie z regułami bycia resortem wiodącym i współpracującym nie jest w stanie wskazać resortów-adresatów.



Rysunek 3. Kody skojarzone

Ze względu na możliwą dynamikę zarówno interpretacji poszczególnych kodów tematycznych jak i zainteresowań resortów należałoby uwzględnić proces uczenia się systemu polegający na modyfikacji ontologii w zakresie powiązań kodów z terminami kluczowymi i pokrewieństwa terminów kluczowych. Reguły uczenia można naszkicować następująco:

- jeśli dokumenty przesyłane jako skojarzone ze względu na pewną parę kodów skojarzonych, znajdują akceptację adresatów, to należałoby te kody opisać dodatkowo terminami kluczowymi, które decydowały o ich skojarzeniu (co uczyni kody powiązanymi),
- jeśli dokumenty przesyłane jako powiązane ze względu na pewną parę kodów powiązanych, znajdują akceptację adresata, to należałoby te kody dopisać do klasy *Kody4[nazwa resortu]-Wsp* definiującej preferencje tego adresata (czyli uczynić kod współpracującym dla tego adresata).

Proces ten nie dotyczy kodów wiodących, gdyż wymagana jest urzędowa decyzja w sprawie wyboru resortów wiodących dla określonych kodów tematycznych.

Klasa *DOKUMENT* reprezentuje pojęcie dokumentu w systemie *U32Mail*. Każdy dokument przeznaczony do zaklasyfikowania tworzony jest w systemie jako instancja tej klasy. Jej podklasy reprezentują preferencje poszczególnych resortów co do interesujących je dokumentów. Wyróżnić można cztery ich grupy:

- *Dok4[nazwa resortu]-Wiodacy* – klasy reprezentujące dokumenty posiadające kody zdefiniowane przez dany resort jako wiodące,
- *Dok4[nazwa resortu]-Wspolpracujacy* – klasy reprezentujące dokumenty posiadające kody zdefiniowane przez dany resort jako współpracujące,

- Dok4[nazwa resortu]-Powiazany – klasy reprezentujące dokumenty powiazane z danym resortem,
- Dok4[nazwa resortu]-Skojarzony – klasy reprezentujące dokumenty skojarzone z danym resortem.

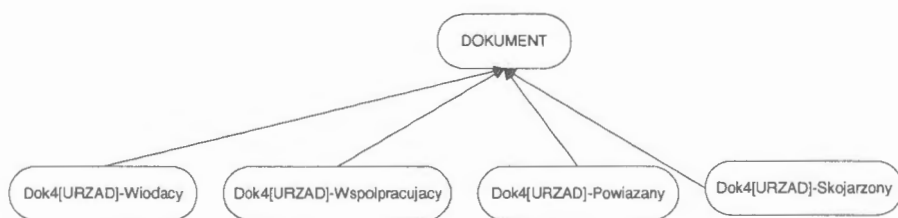
Definicje klas z poszczególnych grup można zilustrować na przykładach dotyczących resortu ARR (Agencja Rynku Rolnego). Klasa Dok4ARR-Wiodacy, w formalizmie logiki deskryptywnej stosowanym przez Protege, zdefiniowana jest za pomocą tak zwanego konstruktora kwantyfikacji szczegółowej: $\exists maKod Kody4ARR-Wiod$. Konstrukcja ta opisuje klasę obiektów, które mają właściwość *maKod* o wartości będącej obiektem należącym do klasy *Kody4ARR-Wiod*. Takie obiekty muszą należeć do klasy *Dokument* (co wynika z definicji właściwości *maKod*) i wchodzić w krąg zainteresowania Agencji Rynku Rolnego, bo taka jest semantyka klasy *Kody4ARR-Wiod*: reprezentuje ona kody tematyczne explicite wskazane przez ARR jako „wiodące”.

Klasa Dok4ARR-Współpracujący zdefiniowana jest za pomocą konstruktora kwantyfikacji szczegółowej: $\exists maKod Kody4ARR-Wsp$, co ma interpretację analogiczną do poprzedniej klasy.

Klasa Dok4ARR-Powiązany zdefiniowana jest za pomocą zagnieżdżonych konstruktorów kwantyfikacji szczegółowej: $\exists maKod (\exists dotyczy (\exists opisujeKod Kody4ARR-Wsp))$.

Klasa Dok4ARR-Skojarzony zdefiniowana jest za pomocą zagnieżdżonych konstruktorów kwantyfikacji szczegółowej: $\exists maKod (\exists dotyczy (\exists pokrewny (\exists opisujeKod Kody4ARR-Wsp)))$

Wstępnie przyjęto więc prostą hierarchię klas zilustrowaną na rys. 4.

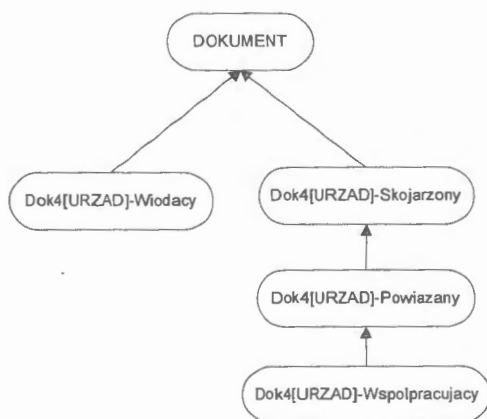


Rysunek 4. Wstępna hierarchia klas typu Dokument

6. Przebieg eksperymentu

Początkowo wszystkie podklasy klas KOD oraz DOKUMENT zostały umieszczone bezpośrednio pod nimi w hierarchii (por. rys. 4). Weryfikacja ontologii przy pomocy serwera RACER wykazała istnienie innej hierarchii, zilustrowanej na rys. 5. Otrzymana hierarchia pozwala przeanalizować zależności pomiędzy

preferencjami wyrażonymi przez poszczególne urzędy co do interesujących je dokumentów. W wypadku przyjętej tu prostej reprezentacji tych preferencji (sprowadzającej się do „alternatywy/dyzjunkcji kodów tematycznych”), analiza taka jest dość trywialna. Zakładając możliwość wyrażenia preferencji poprzez złożone kombinacje kodów tematycznych (włącznie z, np., „zanegowanymi kodami”, czyli wskazaniem, że dokumenty interesujące daną instytucję nie mogą być oznaczone danym kodem), taka analiza może być bardzo interesująca dla projektanta systemu dystrybucji dokumentów.



Rysunek 5. Zweryfikowana hierarchia klas typu Dokument

System wykrył również zależności pomiędzy klasami odnoszącymi się do różnych resortów. Jest to konsekwencją zależności pomiędzy klasami Kody4[nazwa resortu]-Wsp. Uzyskane wyniki związane z analizą hierarchii klas w ontologii pozwalają więc wyciągnąć również wnioski dotyczące współzależności kompetencyjnej poszczególnych resortów. W obecnym systemie, w którym uwzględnione są preferencje wyłącznie urzędów centralnych, jest to zjawisko marginalne i wskazujące jedynie na współdzielenie kompetencji poszczególnych resortów. W przyszłości, gdy zgodnie z planami do systemu dołączone zostaną wszystkie jednostki administracji publicznej, a nawet część samorządowych – analiza tego rodzaju umożliwi w znacznym stopniu skrócenie i zoptymalizowanie ścieżek decyzyjnych, a tym samym ścieżek obiegu dokumentów. Dynamika struktur jednostek organizacji administracyjnej oraz samorządowej oraz fluktuacja kadry (a więc również specjalistów) powoduje, że konieczne jest ciągle wykonywanie tego rodzaju analiz oraz stałe przeorganizowywanie procesów decyzyjnych. Typowe narzędzia wymagają ciągłego nadzoru ekspertów zorientowanych we wszystkich zawiłościach oraz polskiej specyfice środowiska polityczno-ustawodawczego. Proponowane narzędzie wymaga wyłącznie stałego aktualizowania danych związanych z preferencjami tematycznymi poszczególnych jednostek, oraz pewnego

ich priorytetowania (poprzez nadawanie resortom statusu „wiodący” i „współpracujący”).

Po przeprowadzeniu weryfikacji poprawności i spójności systemu ontologicznego, utworzone zostały na potrzeby eksperymentu instancje klasy DOKUMENT odpowiadające dokumentom, które zostały poddane przyporządkowaniu do poszczególnych urzędów. Dokumenty do dystrybucji dobrano tak, aby ilustrowały one wszystkie kategorie: wiodące, współpracujące, powiązane i skojarzone dla wybranego urzędu (Ministerstwo Ochrony Środowiska, MOŚ): Dokumenty te mają następującą postać (właściwości):

- Dok01** maKod(FOREST) – *posiada kod wiodący dla MOŚ*
- Dok02** maKod(ATO, ENV) – *posiada dwa kody przypisane jako wiodące do MOŚ*
- Dok03** maKod(AGRIFORET, AVIATION) – *posiada jeden kod wiodący dla MOŚ i drugi kod, który nie jest ani powiązany ani skojarzonym dla MOŚ*
- Dok04** maKod(DEVGEN) – *posiada kod współpracujący dla MOŚ*
- Dok05** maKod(EDUC, EEE) – *posiada dwa kody współpracujące dla MOŚ*
- Dok06** maKod(FSTR, JUR) – *posiada kod współpracujący dla MOŚ i drugi kod, który nie jest ani powiązany ani skojarzony dla MOŚ*
- Dok07** maKod(FOREST, DEVGEN) – *posiada kod wiodący i kod współpracujący dla MOŚ*
- Dok08** maKod(CULT) – *posiada kod powiązany dla MOŚ (opisany terminem kluczowym „NAUKA I ROZWÓJ”, opisującym również kod DEVGEN, współpracujący dla MOŚ)*
- Dok09** maKod(AUDIO) – *posiada kod skojarzony dla MOŚ (opisany terminem kluczowym „GOSPODARKA”, który jest terminem pokrewnym do terminu „ŚRODOWISKO”, który z kolei opisuje kod współpracujący dla MOŚ)*
- Dok10** maKod(SPORT) – *posiada kod, w żaden sposób niezwiązany z MOŚ*

W celu dokonania klasyfikacji dokumentów uruchamiamy serwer RACER-a i z programu Protege wykonujemy polecenie „Classify Taxonomy”. Powoduje to, przesłanie ontologii do RACER-a i jej analizę, obejmującą, między innymi, wnioskowanie co do przynależności instancji do poszczególnych klas. Wynikiem klasyfikacji dokumentów jest przypisanie ich do klas w następujący sposób (podajemy najbardziej specyficzną klasę przypisaną do danego dokumentu; każdy dokument należy, oczywiście, również do wszystkich nadklas tej klasy).

KLASA	PRZYPISANE DOKUMENTY
Dok4MOS-Wiodacy	Dokument01, Dokument02, Dokument03, Dokument07
Dok4MOS-Współpracujący	Dokument04, Dokument05, Dokument06
Dok4MOS-Powiązany	Dokument08
Dok4MOS-Skojarzony	Dokument09
Dokument	Dokument10

W powyższym eksperymencie wszystkie dokumenty zostały zaklasyfikowane zgodnie z oczekiwaniami. Należy podkreślić, że przykład ten jest prosty i wynik klasyfikacji jest dość oczywisty. Jednocześnie, stworzony mechanizm klasyfikacji dokumentów ma charakter ogólny i może być zastosowany przy dużo bardziej złożonych powiązaniach pomiędzy używanymi pojęciami (klasami).

7. Zakończenie

Przedstawiona w niniejszej pracy ontologia U32MailOnt jest modelem uproszczonym – pod uwagę wzięto przede wszystkim powiązania pomiędzy kodami tematycznymi zawartymi w dokumentach oraz preferencjami urzędów. Przy budowie pełnego systemu zaleca się wziąć pod uwagę również, między innymi, następujące aspekty:

- status dokumentu (dwa dokumenty posiadające te same kody tematyczne powinny zostać zaklasyfikowane do różnych urzędów, jeśli jeden z nich jest dokumentem nad którym pracuje Grupa Robocza UE, a drugi ma trafić do zatwierdzenia do Rady Unii Europejskiej),
- poziom dostępu do informacji niejawnej użytkownika, do którego zostanie skierowany dokument.

Podczas projektowania rzeczywistego systemu należałoby przewidzieć (co jest całkowicie dopuszczalne w zaproponowanym tu modelu) istnienie bardziej złożonych reguł dystrybucyjnych (zawierających m.in. „negację” pewnych kodów dla określonych urzędów).

Stworzony na potrzeby pracy system ontologiczny umożliwia przeprowadzenie analizy spójności danych i adaptację do zmieniających się warunków poprzez:

- wyszukanie kodów, nieprzypisanych do żadnego urzędu jako wiodące,
- wyszukanie kodów, nie przypisanych do żadnego urzędu jako współpracujące,
- wybranie właściwych urzędów dla nowopojawiających się kodów (m.in. przez powiązanie ich z odpowiednimi terminami kluczowymi)

Należy założyć, że ontologia będzie dostrajana w trakcie eksploatacji systemu. Dodatkową korzyścią z posiadania takiej ontologii jest możliwość stworzenia mapy powiązań kompetencyjnych pomiędzy poszczególnymi resortami.

Literatura

- Bechhofer S., van Harmelen F., Hendler J., Horrocks I., McGuinness D.L., Patel-Schneider P.F., Stein L.A., Olin F.W., *OWL Web Ontology Language Reference*, <http://www.w3.org/TR/2003/PR-owl-ref-20031215/>
- Gómez-Pérez A., *Ontological engineering*, Universidad Politécnica de Madrid, <http://www.ontology.org/main/presentations/madrid/theoretical.pdf>
- Haarslev V., Moller R., *RACER User's Guide and Reference Manual Version 1.7.7*, <http://kogs-www.informatik.uni-hamburg.de/~haarslev/publications/racer-manual-1-6.pdf>
- Haarslev V., Moller R., *Racer: A Core Inference Engine for the Semantic Web*, http://km.aifb.uni-karlsruhe.de/ws/eon2003/EON2003_Haarslev.pdf
- Ławcewicz K.A. (2003) Elektroniczny system przepływu i archiwizowania dokumentów europejskich (EWD). W: Hołubiec J. (Red.): *Analiza systemowa w finansach i zarządzaniu. Wybrane problemy*, 5, IBS PAN, Warszawa.
- Koncepcja wstępna systemu Przepływu i Archiwizacji Spraw i Dokumentów Elektronicznych Rady Unii Europejskiej*, lipiec 2003.
- Noy N.F., McGuinness D.L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*, Stanford Medical Informatics Technical Report SMI-2001-0880.
- Protégé User's Guide*, <http://protege.stanford.edu>
- Ramy instytucjonalne oraz kody tematyczne*, Rada Unii Europejskiej, Sekretariat Generalny, kwiecień 2003 (wersja angielska i polskie tłumaczenie).
- Rules and Procedures applicable to the transmission of Official Documents (via the U32MAIL system)*, Rada Unii Europejskiej, Sekretariat Generalny, styczeń 2003.
- Smith M.K., Welty Ch., McGuinness D.L. (2003) *OWL Web Ontology Language Guide*, <http://www.w3.org/TR/2003/PR-owl-guide-20031215>
- System koordynacji polityki europejskiej w okresie poprzedzającym członkostwo Polski w UE, UKIE*, luty 2003.
- U32Mail – Guide to browsing Council documents - User guide*, Rada Unii Europejskiej, marzec 2003.

IBS PAN *Seria*

45187

Bibl. podręczna

ISSN 0208-8028

ISBN 83-85847-92-8

**W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy
prosimy o kontakt z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 837-35-78 w. 241 e-mail: biblioteka@ibspan.waw.pl**