



**INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK**

**ANALIZA SYSTEMOWA W FINANSACH
I ZARZĄDZANIU**

Wybrane problemy
Tom 11

Pod redakcją
Jerzego HOŁUBCA

Warszawa 2009

Wykaz opiniodawców artykułów zamieszczonych
w niniejszym tomie:

prof. dr hab. inż. Jerzy HOŁUBIEC
dr inż. Lech KRUŚ
doc. dr hab. inż. Wiesław KRAJEWSKI
doc. dr hab. Jacek MALINOWSKI
dr inż. Edward MICHALEWSKI
prof. dr Adam SKOREK
dr hab. Ryszard SMARZEWSKI
prof. dr hab. inż. Andrzej STRASZAK
dr Dominik ŚLĘZAK
prof. dr hab. inż. Stanisław WALUKIEWICZ
doc. dr hab. Sławomir ZADROŻNY

© Instytut Badań Systemowych PAN
Warszawa 2009

ISBN 9788389475220

Druk: Zakład Poligraficzny Jerzy Kosiński, Warszawa

ZASTOSOWANIE ONTOLOGII W ZARZĄDZANIU DOKUMENTAMI ELEKTRONICZNYMI DLA CELÓW ARCHIWIZACJI

Marek Gajewski

Studia Doktoranckie IBS PAN

Celem artykułu jest przedstawienie wybranych spośród obowiązujących standardów i kierunków rozwoju technik informatycznych w zakresie zarządzania dokumentem elektronicznym, ze szczególnym uwzględnieniem zastosowań w dziedzinie archiwistyki. Centralnym punktem zainteresowania są metody oparte na semantycznej analizie dokumentów, w szczególności z zastosowaniem reprezentacji wiedzy w postaci ontologii. W pracy omawia się wybrane przykłady tego typu podejść i przedstawia się wizję własnych prac w tym zakresie.

Słowa kluczowe: dokument elektroniczny, archiwizacja, ontologia.

Wstęp

Początków myśli ludzkiej dotyczącej zarządzania dokumentami można się doszukiwać już ponad 5000 lat temu [16]. Sumerowie i Egipcjanie tworzyli kolekcje dokumentów zapisanych na tabliczkach pismem klinowym oraz na papirusie za pomocą skryptu hieroglificznego. Zawarte informacje dotyczyły prawnych i ekonomicznych transakcji takich, jak: umowy, akty własności czy oświadczenia woli.

W dobie rozwoju Społeczeństwa Informacyjnego liczba tworzonych i przetwarzanych dokumentów rośnie lawinowo. Zarządzanie nimi bez wsparcia ze strony systemów informatycznych staje się niemożliwe lub, co najmniej, niepraktyczne, nawet w przypadku małych firm i instytucji.

Dokumenty różnego rodzaju stanowią ważny element dokumentacji zarówno działalności firm, jak i dokonań poszczególnych osób. Z tego względu stanowią one zazwyczaj przedmiot szczegółowych regulacji prawnych, dotyczących warunków ich przechowywania, opisywania itd. Zagadnienia te stanowią istotę *archiwistyki*, dziedziny ulegającej obecnie ogromnym zmianom związanym z koniecznością uwzględnienia rosnącej roli mediów elektronicznych i ich specyfiki.

W niniejszej pracy dokonujemy krótkiego przeglądu zagadnień istotnych dla sprawnej archiwizacji rosnących zasobów dokumentów elektronicznych. Możliwości podniesienia efektywności realizowanych w tym względzie procesów upatrujemy w zastosowaniu jawnej reprezentacji wiedzy w postaci *ontologii*. Ontologie już od wielu lat stosowane są przy przetwarzaniu dokumentów tekstowych, w mniejszym lub większym zakresie. Dopiero jednak ostatnio pojawiły się standardy i narzędzia umożliwiające zastosowanie ich na szerszą skalę. W pracy przedstawiamy wybrane, związane z tym, zagadnienia i przykłady zastosowań znane z literatury.

1. Pojęcie dokumentu

Pojęcie dokumentu może być szeroko rozumiane. W potocznej świadomości utożsamiane jest z pismem urzędowym, rodzajem filmu, przedmiotem świadectwa kulturowego, plikiem tekstowymi itp. Źródła bibliograficzne definiują sam termin w następujący sposób:

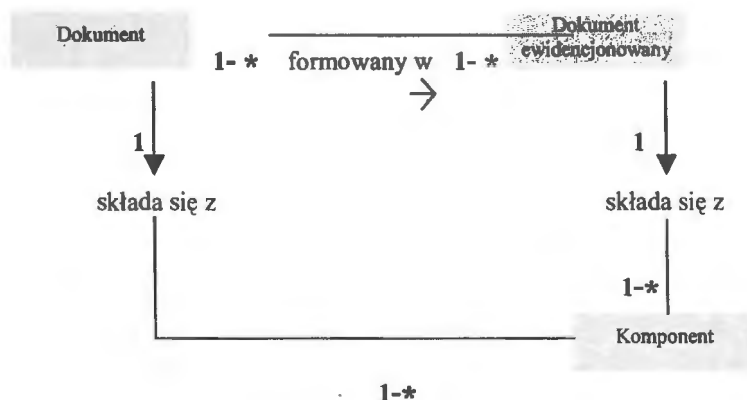
- każdy przedmiot materialny będący świadectwem jakiegoś faktu, zjawiska, lub myśli ludzkiej(...) /Encyklopedia PWN/[19]
- zapisana informacja lub obiekt, który może być traktowany jako jednostka /ISO 15489/[8]

W niniejszej pracy przyjmujemy terminologię stosowaną w specyfikacji MoReq2 [12], opracowanej na zamówienie Unii Europejskiej. Specyfikacja ta obejmuje szerokie spektrum zagadnień interesujących z punktu widzenia archiwizacji. Wprowadza ona wiele pojęć szczegółowych. Ich polskie odpowiedniki określają różne krajowe regulacje prawne, w szczególności rozporządzenie MSWiA [14]. Powszechne przyjęcie standardu określonego specyfikacją MoReq2 przez kraje członkowskie stworzy warunki do rozwoju stosownych narzędzi informacyjnych, zapewniając im zbyt na dużym rynku europejskim.

Wiodącą rolę w pracach nad standardem MoReq2 odgrywa organizacja DLM-Forum, powstała z inicjatywy Unii Europejskiej, skupiająca archiwa państwowe krajów członkowskich oraz innych partnerów zainteresowanych archiwizacją, zarządzaniem dokumentami itp. Zaangażowanie Unii Europejskiej najlepiej wyrażają zalecenia w *sprawie działań priorytetowych w celu wzmocnienia współpracy w dziedzinie archiwów w Europie* [20]. Wykonanie specyfikacji zostało zlecone firmie Serco Consulting w ramach projektu Unii Europejskiej IDABC.¹ W bieżące prace zaangażowani byli również eksperci zewnętrzni. Ważną rolę ode-

¹ Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens.

grały zespoły wolontariuszy, którzy na bieżąco poddawali przeglądowi i ocenie elementy tworzonego standardu.



Rys. 1. Reprezentacja dokumentu w systemie informatycznym.
Opracowanie: Na podstawie specyfikacji MoReq2 [12]

Specyfikacja MoReq2² określa dokument w środowisku informatycznym (Rys.1) jako złożony z jednego lub większej liczby komponentów. Przez komponent należy rozumieć *plik komputerowy*. Wewnątrz pliku komputerowego wystąpić może odwołanie do innego pliku i dopiero łącznie traktowane one będą wtedy jako dokument.

Centralne pojęcie specyfikacji MoReq2 wyraża się angielskim terminem *record*, który zgodnie z rozporządzeniem [15] należy tłumaczyć na język polski jako *dokument ewidencjonowany*. Dokumenty ewidencjonowane to „dokumenty elektroniczne świadczące o wykonywaniu działalności podmiotów, powstające w nich lub napływające do nich”, które „odzwierciedlają przebieg załatwiania i rozstrzygania spraw” i w związku z tym „podlegają ewidencjonowaniu w systemie teleinformatycznym” [15]. Szczególną ich cechą jest ich finalny charakter – przyjmuje się, że dokument taki nie podlega już żadnym dalszym zmianom.

W niniejszej pracy interesują nas przede wszystkim dokumenty tekstowe. Kwestia zarządzania nimi w systemie informatycznym jest bardzo istotna. Obecne rozwiązania biznesowe w dużym stopniu nadal opierają się na traktowaniu dokumentu jako czarnej skrzynki z punktu widzenia systemu informatycznego. Z dru-

² Model Requirements for the management of electronic records. [12]

giej strony istnieje wiele prac podejmujących temat zarządzania dokumentami z zastosowaniem metod automatycznej interpretacji ich zawartości; por. np. [2,5].

2. Standardy opisu metadanymi i klasyfikacji dokumentów

Pojęcie metadanych staje się coraz popularniejsze w dobie rozwoju i wszechobecności systemów informatycznych, w szczególności baz danych. Dane o danych, a w interesującym nas przypadku danc o treści dokumentu tekstowego, stanowią znaczenie przedmiotowego terminu. Treść dokumentu tworzona jest w języku naturalnym z przeznaczeniem do odczytu przez człowieka. Nie opracowano dotychczas skutecznej metody pełnej semantycznej interpretacji języka naturalnego przez komputer. Dla umożliwienia zarządzania plikiem tekstowym tworzy się z użyciem metadanych jego reprezentację, która pozwala na dokonywanie operacji logicznych celem odszukania i dalszego opracowania na dokumencie.

Dla efektywnego zastosowania metadanych istotne jest powszechne przyjęcie określonego standardu ich postaci. Można wskazać dwie drogi prowadzące do przyjęcia standardu. Pierwsza polega na stosowaniu, na zasadzie dobrowolnej akceptacji, norm wypracowanych przez pozarządowe organizacje takie, jak ISO, DLM Forum, Dublin Core Metadata Initiative (DCMI), OASIS. Druga polega na narzuceniu standardu przez akty normatywne.

W warunkach polskich jednym z ważniejszych aktów normatywnych w tej dziedzinie jest Rozporządzenie MSWiA w sprawie niezbędnych elementów struktury dokumentów elektronicznych [14]. Wyszczególnia ono metadane wymagane przy opisie dokumentów elektronicznych powstałych w organach państwowych i państwowych jednostkach organizacyjnych, w organach jednostek samorządu terytorialnego i samorządowych jednostkach organizacyjnych. Łatwo dostrzec w tym dokumencie wzorowanie się na standardzie Dublin Core Metadata Element Set [3,9].

Z użyciem informacji o dokumentach, której dostarczają metadane, można przeprowadzić ich klasyfikację (pogrupowanie). Można zastosować grupowanie uwzględniające związek poszczególnych dokumentów z różnymi procesami biznesowymi realizowanymi w danej firmie/organizacji, czyli przynależność dokumentów do różnych akt sprawy. Następnie, można dokonać grupowania ze względu na aspekt merytoryczny (grupa dokumentów), lub techniczny (tomy).

Dokumentacji można nadać układ rzeczowy. Jest to rodzaj klasyfikacji polegający na przypisaniu akt do grup w układzie hierarchicznym, reprezentującym całość aktywności danego podmiotu. Przykład, określający jednocześnie standard przedmiotowej klasyfikacji w prawodawstwie polskim, znajduje się w Rozporządzeniu *w sprawie postępowania z dokumentacją, zasad jej klasyfikowania i kwali-*

fikowania oraz zasad i trybu przekazywania materiałów archiwalnych do archiwów państwowych [13]. We wzorcu przyjęto występowanie w każdej jednostce organizacyjnej czterech głównych klas rzeczowych: zarządzanie, kadry, środki rzeczowe, ekonomika. Dookreślenie kolejnych klas głównych pozostaje w gestii ustaleń wewnętrznych.

Element	Dublin Core	Rozporządzenie
Współtwórca	✓	
Zakres	✓	
Twórca	✓	✓
Data	✓	✓
Opis	✓	✓
Format	✓	✓
Identyfikator	✓	✓
Język	✓	✓
Wydawca	✓	
Relacja	✓	✓
Prawa (Uprawnienia)	✓	✓
Źródło	✓	
Temat	✓	
Tytuł	✓	✓
Typ	✓	✓
Dostęp		✓
Odbiorca		✓
Grupowanie		✓
Kwalifikacja		✓

Tabela 1 Zestawienie elementów metadanych określonych w Dublin Core [9] i wymienionych w polskim prawodawstwie (Dz. U. nr 206 poz. 1517 z 2006r) [14]

3. Ontologie i ich rola w zarządzaniu dokumentem – przykłady zastosowań

Głównym przedmiotem naszego zainteresowania w niniejszej pracy są możliwości zastosowania ontologii w systemach zarządzania dokumentami. Człowiek jest w stanie analizować treść dokumentu i na tej podstawie zarządzać nim w odpowiedni sposób. Ograniczone możliwości zastosowania w ten sposób narzędzi komputerowych wynikają z trudności automatycznej analizy semantycznej dokumentu. Człowiek potrafi odczytać tekst ze zrozumieniem jego treści między innymi

dzięki odpowiedniemu aparatowi pojęciowemu, związanemu z posiadaną przez niego wiedzą. Istotne jest, że znaczenie danego pojęcia jest określane w danym kontekście przez inne pojęcia. Stanowi to punkt wyjścia dla rozpatrywania metod formalnej reprezentacji wiedzy, pozwalających na jej maszynowe wykorzystanie. Ważną rolę odgrywa tu pojęcie ontologii.

Różnych szczegółowych definicji ontologii jest tyle ilu autorów, którzy zajmują się tą tematyką. Wszystkie one wydają się mieć jednak wspólny mianownik, który dobrze wyraża zwięzła definicja podana przez Grubera [6]: *ontologia* stanowi jawną specyfikację konceptualizacji pewnego wycinka wiedzy, przy czym pod pojęciem konceptualizacji rozumie się zbiór obiektów, pojęć i innych bytów występujących w pewnym obszarze zainteresowania oraz powiązań pomiędzy nimi. Konceptualizacja stanowi pewien abstrakcyjny, uproszczony model świata, który chcemy reprezentować w pewnym określonym celu.

Pojęcie ontologii stosowane jest w odniesieniu do szerokiego spektrum obiektów. Ontologie mogą mieć postać mniej lub bardziej formalną. Na przykład, ontologią można nazwać tezaurus, katalog, układ rzeczowy klasyfikacji dokumentów itp. Istota użyteczności takich narzędzi polega na pewnym ograniczeniu liczby i uporządkowaniu zbioru stosowanych haseł/pojęć/kategorii, z ewentualnym umieszczeniem ich w układzie hierarchicznym. Można to uznać za próbę standaryzacji wiedzy. Inny rodzaj ontologii może zawierać sformalizowane struktury, porządkujące pojęcia w sposób spójny z ich znaczeniem. To „znaczenie” stanowi odzwierciedlenie interpretowania danego pojęcia przez człowieka w świecie rzeczywistym. Są to opisane formalnie taksonomie, własności, relacje i ograniczenia pozwalające zidentyfikować dane pojęcie w grupie innych pojęć dla rozpatrywanej dziedziny.

Podstawowe elementy składające się na ontologię to *klasy*, *własności* i *instancje*. Opiszemy je teraz pokrótce.

Klasa obejmuje zbiór obiektów z dziedziny, której dotyczy ontologia, wyróżniony w tej dziedzinie, jako reprezentujący pewne pojęcie. Pojedynczy reprezentant klasy jest nazywany jej *instancją*. Poszczególne klasy tworzą hierarchię, która stanowi taksonomiczną część ontologii. Jeśli klasa *k1* występuje w tej hierarchii poniżej klasy *k2*, to oznacza to, że wszystkie instancje klasy *k1* są również instancjami klasy *k2*.

Własności określają relacje zachodzące pomiędzy obiektami modelowanej dziedziny lub pomiędzy obiektami i wartościami standardowych typów danych takich, jak liczby całkowite czy łańcuchy znaków. Zazwyczaj dopuszcza się możliwość określenia, jakich klas instancje mogą pozostawać w relacji określonej daną wła-

ściwością. Na przykład, właściwość `jestAutorem` może wiązać w pewnej ontologii wyłącznie instancje klasy `Osoba` z instancjami klasy `Dokument`.

Ontologie pozwalają rozszerzyć użyteczność metadanych, wpisując je w sformalizowaną strukturę pojęciową, wyrażającą wiedzę z danej dziedziny. Wiedza ta, dzięki takiemu formalnemu wyrażeniu w postaci ontologii, staje się interpretowalną przez komputer.

Ontologie stanowią pomoc dla człowieka opatrującego dokument metadanymi. Jednak ostatecznym celem ich zastosowania jest stworzenie oprogramowania zdolnego do automatycznego lub półautomatycznego (w interakcji z użytkownikiem) tworzenia takich opisów. Jak zauważają Blythe i Gil [4], ręczne tworzenie sformalizowanego opisu (ang. *annotation*) może być właściwsze od automatycznego, jednakże niesie ze sobą pierwiastek subiektywnej analizy tekstu. Można przyjąć, że dla danego dokumentu, z użyciem tej samej ontologii, można otrzymać różne opisy, zależnie od osoby przygotowującej opis. Zaproponowane w artykule [4] rozwiązanie przedstawia przykład użycia ontologii do opisu dokumentów dotyczących rozgrywek piłkarskich. Narzędzie ACE (ang. *Annotation Canonicalization through Expression synthesis*) umożliwia normalizację opisu dokumentu utworzonego przez użytkownika w formie swobodnego tekstu w języku naturalnym. Na podstawie podanego opisu system automatycznie generuje zbiór jego zmodyfikowanych form, w ten sposób, że mają one jednoznaczną interpretację w terminach rozważanej ontologii. Użytkownik wybiera spośród zaproponowanych form tę, która najlepiej według niego oddaje jego oryginalny opis lub w trybie interakcyjnym poszukuje lepszego sformułowania, ewentualnie rozszerzając przy tym ontologię. Tak więc, rolą ontologii jest tu dostarczenie bazy ustandaryzowanych pojęć, które mogą doprecyzować opis dokumentu, utworzony przez użytkownika w języku naturalnym. Tworzy się tym samym system koherentnej terminologii dla celów komunikacji pomiędzy użytkownikami zainteresowanymi tekstami z danej dziedziny sportu.

Ontologie mają zastosowanie również we wsparciu procesu wyodrębniania informacji z dokumentów. W tym kontekście warto wspomnieć narzędzie MnM [18]. Pozwala ono automatycznie indeksować dokumenty tekstowe z użyciem znaczników (ang. *tag*) SGML/XML. Mechanizm działania polega na uprzednim nauczeniu narzędzia przeznaczonego do ekstrakcji informacji (ang. *Information Extraction, IE*) na pewnej liczbie dokumentów uczących. Użytkownik ręcznie oznacza w nich wybrane informacje, z użyciem znaczników odpowiadających pojęciom z wybranej ontologii (ang. *ontology driven mark-up*). Na podstawie dokumentów uczących konstruowane są reguły ekstrakowania poszukiwanych informacji.

W literaturze zaproponowano bardzo wiele formalizmów reprezentacji wiedzy dziedzinowej, w tym również ontologii. W związku z tym znaczenia nabiera zagadnienie dopasowywania do siebie ontologii [11]. Przykład rozwiązania pozwalającego na integrację dwóch różnych systemów reprezentacji wiedzy w kontekście opisywania dokumentów przedstawia praca [1]. Celem jest automatyczne ustalenie odpowiedniości znaczników stosowanych przez moduł analizy językowej tekstu z pojęciami występującymi w przyjętej ontologii wiedzy dziedzinowej. Zadanie to realizowane jest poprzez ręczne sformułowanie reguł określających taką odpowiedniość (ang. *knowledge acquisition rules*). Reguły wyrażane są w języku XPath i wiązane są z pojęciami ontologii dziedzinowej, których dotyczą. Pozwala to na stworzenie instrumentu do automatyczne wzbogacenie wyników analizy językowej treści dokumentów o pojęcia zawarte w ontologii.

Kolejny przykład zastosowania ontologii do zarządzania dokumentami, a w szczególności aktami notarialnymi, zawiera praca [2]. Przedstawia ona narzędzie wyodrębniające z dokumentu swoiste informacje, relewantne dla zaindeksowania dokumentu w systemie informacyjnym. Proces, mający na celu osiągnięcie założonego celu wykorzystuje trzy typy ontologii. *Ontologia struktur dokumentów* jest wykorzystywana na etapie rozbioru dokumentu na funkcjonalne fragmenty. Na przykład, może ona opisywać strukturę notarialnego aktu sprzedaży nieruchomości, wyróżniając w nim takie fragmenty, jak wstęp, część opisującą kontrahentów i część opisującą nieruchomość. W ogólności wynik takiego rozbioru może nie być jednoznaczny i spośród uzyskanych podziałów wybiera się jeden, który najlepiej pasuje do zbioru słów kluczowych, nazw własnych itd. dodatkowo przypisanych każdej z wyróżnionych części. W ten sposób podzielony dokument poddawany jest analizie lingwistycznej i semantycznej z wykorzystaniem odpowiednio *ontologii leksykalnej* i *dziedzinowej*. Ontologia dziedzinowa służy finalnie do utworzenia reprezentacji dokumentu z użyciem języka RDF.

Wykorzystanie ontologii do zarządzania dokumentem umożliwia automatyczne przetwarzanie informacji, a dzięki sformalizowanej strukturze także jej wymianę pomiędzy systemami.

Podsumowanie

W rozważanym przez nas kontekście zarządzania dokumentami, ontologię można traktować, jako swego rodzaju umowę pomiędzy ludźmi, że w danej dziedzinie, przy zastosowaniu komputera do przetwarzania dokumentów, będziemy posługiwać się ograniczoną listą pojęć do reprezentacji dokumentów. Uzyskuje się w ten sposób pewien quasi-naturalny język, zapewniający większą jednoznaczność komunikacji i możliwość automatycznej, komputerowej interpretacji jego wyrażań. Ze względu na złożoność języka naturalnego nie zostało dotychczas opracowane

narzędzie informatyczne zdolne do pełnego wykorzystania tego medium w komunikacji z użytkownikiem.

Z punktu widzenia archiwistyki, zbiór dokumentów jest odzwierciedleniem przebiegu prowadzonej sprawy w miejscu zaistnienia konieczności poświadczenia wykonanych czynności. Można powiedzieć, że dokumenty stanowią bazę wiedzy na temat działalności danego podmiotu. Wiedza, aby mogła być przetwarzana komputerowo musi być w odpowiedni sposób skodyfikowana. Wybór metody kodyfikacji determinuje zakres przetwarzania komputerowego pozyskanych danych oraz operacji na źródle danych. Ontologie, ze względu na swoją popularność i teoretyczne podstawy, stanowią obiecujące narzędzie w tym względzie.

W zadaniach archiwizacyjnych koncentrujemy się na takim opracowaniu dokumentu, aby był on właściwie zidentyfikowany merytorycznie oraz zaklasyfikowany według stosownych kryteriów przechowywania. Przez identyfikację merytoryczną należy rozumieć rozpoznanie treści dokumentu oraz jego powiązań w ramach prowadzonych spraw. Rodzaj dokumentu, może być identyfikowany przez reguły dopasowania swoistych danych z wykorzystaniem ontologii dziedzinowej. Sprawy, w kontekście zarządzania dokumentami, można zidentyfikować jako obiekty należące do grupy procesów. Zachodzi w tym przypadku możliwość ograniczenia ich do zbioru spraw właściwych danemu podmiotowi.

Spełnienie wymagań regulacji prawnych dotyczących archiwizacji będzie wspomagane poprzez możliwość ich wyrażenia w postaci ontologii archiwizacji i zastosowania automatycznego procesu wnioskowania.

W systemie zarządzania dokumentami pożądane jest uwzględnienie ścieżki załatwiania danej sprawy, związanej z pojęciem obiegu dokumentu (ang. *work-flow*). Szczególnie ważne są punkty decyzyjne w procesie załatwiania oznaczonej sprawy. Tworzą one alternatywne rozwiązania przebiegu danej sprawy. Implikuje to konieczność utworzenia odpowiedniego mechanizmu w systemie informatycznym, zarządzającego wariantami obiegu dokumentu. Możemy utworzyć pewne reguły oparte na aksjomatach, „tłumaczące” komputerowi, co to jest odwołanie, awans (związane z tym uaktualnienie akt zarejestrowanych dla sprawy zatrudnienie), brak rozstrzygnięcia przetargu itp. Należy także w badaniach zwrócić uwagę na możliwość wykorzystania dokumentów z zakresu prawodawstwa polskiego do walidacji ontologii dziedzinowej z przepisami normującymi prowadzenie spraw i archiwizowanie dokumentów.

Opracowanie skutecznej metodologii pełnego semantycznego rozpoznania treści dokumentu na użytek procesu archiwizacji wymaga jeszcze rozwiązania wielu problemów [5]. Zastosowanie ontologii już doprowadziło do opracowania wielu interesujących podejść i wydaje się stanowić bardzo interesujący kierunek dalszych prac w tej dziedzinie.

Literatura

- [1]. Amardeilh F., Laublet P., Minel J.-L (2005): Document Annotation and Ontology Population from Linguistic Extractions. Proceedings of Knowledge Capture (KCAP05).
- [2]. Amato F., Mazzeo A., Penta A., Picariello A. (2008): Using NLP and Ontologies for Notary Document Management Systems. 19th International Workshop on Database and Expert Systems Applications (DEXA 2008).
- [3]. ANSI/NISO Z39.85-2001 - The Dublin Core Metadata Element Set.
- [4]. Blythe J., Gil Y. (2004): Incremental Formalization of Document Annotations through Ontology-Based Paraphrasing. Proceedings of the 13th International Conference on World Wide Web, New York.
- [5]. Corcho O. (2006): Ontology based document annotation: trends and open research problems Int. J. Metadata, Semantics and Ontologies. Vol.1, No. 1, 47-57.
- [6]. Gruber Th. R. (1993): A Translation Approach to Portable Ontology Specification Knowledge Systems Laboratory Technical Report KSL 92-71.
- [7]. Harris P., Wibel S. L. (2004): Metadata: a worldwide library effort. ISO Focus April, 31-32.
- [8]. ISO 15489:2001. Information and documentation: Records management.
- [9]. ISO 15836:2003. Information and documentation – The Dublin Core metadata element set.
- [10]. ISO/TR 15489-2. Information and documentation: Records management.
- [11]. Kalfoglou Y., Schorlemmer M. (2005): Ontology mapping: the state of the art. <http://drops.dagstuhl.de/opus/volltexte/2005/40/>
- [12]. Model Requirements for the Management of Electronic Records, update and extension (2008): <http://www.moreq2.eu/downloads.htm>
- [13]. Rozporządzenie Ministra Kultury w sprawie postępowania z dokumentacją, zasad jej klasyfikowania i kwalifikowania oraz zasad i trybu przekazywania materiałów archiwalnych do archiwów państwowych .Dz. U. nr 167 poz. 1375, 2002.
- [14]. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji w sprawie niezbędnych elementów struktury dokumentów elektronicznych. Dz. U. nr 206 poz. 1517, 2006.
- [15]. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji w sprawie szczegółowego sposobu postępowania z dokumentami elektronicznymi. Dz. U. nr 206 poz. 1518, 2006.
- [16]. The New Encyclopædia Britannica, Vol. 21 Macropædia, Chicago, 1993, 615-629.
- [17]. Ustawa o narodowym zasobie archiwalnym i archiwach. Dz. U. nr 97 poz. 673, 2006.

- [18]. Vargas-Vera M., Motta E., Domingue J., Lanzoni M., Stutt A., Ciravegna F. MnM: Ontology Driven Semi-Automatic and Automatic Support for Semantic Markup. Proceedings The 13th International Conference on Knowledge Engineering and Management (EKAW 2002), 379-391.
- [19]. Wielka Encyklopedia PWN, tom 7. Wyd. Nauk. PWN, Warszawa 2002.
- [20]. Zalecenie Rady z dnia 14 listopada 2005 r. w sprawie działań priorytetowych w celu wzmocnienia współpracy w dziedzinie archiwów w Europie. Dziennik Urzędowy UE L 312 z 2005r.

ISBN 9788389475220