

XML i nowoczesne metody zarządzania treścią

Wykład 12: Zarządzanie dokumentami

Maciej Ogrodniczuk

MIMUW, 2 stycznia 2011

Dlaczego zarządzanie dokumentami jest ważne?

Bo dokumenty są ważne.

90% zasobów informacyjnych firm jest przechowywanych w dokumentach, a nie w bazach danych.

(Deloitte and Touche)

Rodzaje i odmiany systemów do zarządzania dokumentami:

- Web Content Management System – do zarządzania zawartością witryny internetowej,
- Enterprise Content Management System – zarządzanie dokumentami biznesowymi organizacji,
- system obiegu dokumentów kancelaryjnych,
- system publikacyjny,
- portal korporacyjny,
- system do pracy grupowej,
- elektroniczne archiwum,
- ...

„Tradycyjne” metody zarządzania dokumentami:

- obieg dokumentów papierowych (szafy, segregatory, asystentka, goniec...),
- poczta elektroniczna,
- dyskietki (brrr!),
- pen-drive'y,
- dyski sieciowe,
- ...

Problemy (i rodzące się potrzeby):

- redundancja (ta sama informacja powtórzona w wielu miejscach) wobec możliwości wielokrotnego wykorzystania (ang. *reuse*),
- nieaktualne informacje,
- kłopoty ze znalezieniem właściwej informacji,
- trudności z koordynacją prac zespołu autorów/redaktorów,
- brak możliwości publikacji (i to wielomedialnej),
- brak personalizacji.

Jeszcze więcej problemów z dokumentami

Co robić, gdy dokumenty mają:

- dużą objętość?
- spory stopień komplikacji?
- wysoką wartość?
- długi cykl życia?
- dużą częstość aktualizacji?

a ponadto funkcjonują w organizacji:

- rozproszonej geograficznie,
- wieloosobowej,
- o dużej specjalizacji?

Rozwiązaniem są:

- systemy zarządzania dokumentami (CMS/DMS, ang. *Content/Document Management Systems*),
- systemy wyszukiwawcze (IA, ang. *Information Access*),
- systemy zarządzania wiedzą (KMS)

(albo ich prostsze odpowiedniki).

Typowe funkcje:

- centralne składowanie dokumentów,
- lokalne kopie, synchronizowane z repozytorium,
- blokowanie dokumentów do edycji i zwalnianie blokady po edycji,
- wersjonowanie dokumentów,
- możliwość równoległej edycji dokumentów przez wiele osób i scalanie dokumentów.

Cała masa rozwiązań:

CVS (Concurrent Versions System), SVN (Subversion), IBM Rational ClearCase, Bazaar, BitKeeper, Git...

Czyli strony internetowe edytowalne „przez każdego”.

Cechy charakterystyczne:

- każdy użytkownik ma prawo do edycji,
- rozwiązanie działa bez dodatkowych wtyczek, bezpośrednio w przeglądarce,
- do tworzenia treści wykorzystuje się uproszczoną składnię znacznikową.

Przedstawiciele: MediaWiki, MoinMoin, TiddlyWiki...

- 1 repozytorium – scentralizowana, neutralna pula zasobów,
- 2 aplikacja:
 - funkcjonalność biznesowa,
 - obieg dokumentów, obsługa procesów (ang. *workflow*),
 - wyszukiwanie,
 - prezentacja/publikacja,
- 3 interfejs użytkownika:
 - system nawigacji,
 - system redakcyjny.

Repozytorium:

- dokumentów:
 - możliwość zapisu dowolnych typów dokumentów,
 - wersjonowanie,
 - blokowanie dokumentów do edycji:
 - pesymistyczne – unikamy konfliktów za wszelką cenę, zakładamy blokadę już przy pobieraniu danych do edycji,
 - optymistyczne – konflikty nie są częste, więc możemy chronić samą operację modyfikacji,
 - specjalne wsparcie dla XML-a,
- metadanych (informacji o dokumencie – autorach, dacie publikacji, numerze wersji...):
 - zapis metadanych zwykle poza dokumentami – konieczność synchronizacji,
 - najchętniej: możliwość dowolnej konfiguracji metadanych (nazw, typów, etykiet, warunków prezentacji w interfejsie, warunków synchronizacji, ...)
 - czasem metadane strukturalne (listy, drzewa).

Chodzi o „automatyzację procesów biznesowych, w całości lub w części, podczas której dokumenty, informacje lub zadania są przekazywane od jednego uczestnika do następnego według odpowiednich procedur zarządczych”.

Workflow Management Coalition, www.wfmc.org

Dwa składniki:

- działania osób sterują procesem,
- proces steruje działaniami osób.

Zdefiniowanie procesu to określenie co najmniej:

- kolejnych faz pracy nad dokumentami (stanów, ang. *states*),
- dozwolonych przejść między stanami,
- ról użytkowników uprawnionych do pracy nad dokumentem w danym stanie.

1 zarządzanie treścią:

- wszystkie zasoby dostępne dla (uprawnionych) użytkowników,
- użytkownik decyduje, z których zasobów w danej chwili korzysta,
- typowy sposób dostępu:
 - przeglądanie katalogów,
 - wyszukiwanie,

2 zarządzanie procesem:

- ściśle zdefiniowane role i kompetencje,
- użytkownik wykonuje zadania wskazane przez system,
- system zarządza przekazywaniem dokumentu kolejnym osobom,
- typowy sposób dostępu: lista zadań do wykonania.

Najczęściej spotykane elementy interfejsu:

- struktura drzewiasta
(drzewo katalogów, kategorii, metadanych...)
- lista metadanych,
- dokumenty powiązane,
- wyszukiwarka,
- filtry wyników,
- ...

Konfiguracja:

- typów dokumentów,
- katalogów i ich dopuszczalnej zawartości,
- metadanych,
- typów dowiązań,
- schematu przepływu prac,
- uprawnień,
- pluginów,
- ...

To nie tylko hiperlinki – to w ogólności wszystkie rodzaje powiązań między dokumentami i treścią.

Rodzaje linków:

- OO: między całymi dokumentami,
- CO: do całych dokumentów (hiperlinki, dołączanie dokumentów podrzędnych),
- CC: do fragmentów dokumentów (hiperlinki, zarządzanie wersjami i wariantami),
- jedno- lub dwukierunkowe,
- o dwóch lub więcej końcach (kotwicach),
- opisane metadanymi.

Miejsce przechowywania:

- pełna informacja w dokumencie,
- identyfikatory w dokumencie, pełna informacja w bazie,
- komplet informacji w bazie danych (ze ścieżkami do fragmentów dokumentu).

Cel: możliwość powrotu do wcześniejszej wersji.

Wielopoziomowe tworzenie wersji:

- tworzone automatycznie w momencie zapisu dokumentu (ang. *revisions*):
 - przy każdym zapisie,
 - w momencie zwolnienia blokady,
- wydania tworzone na żądanie (ang. *releases*):
 - w dowolnym (przełomowym) momencie życia dokumentu,
 - w momencie publikacji – wszystkie składniki dokumentu są zwykle „zamrażane”.

Ważne: nie tylko dokumenty, ale także:

- metadane,
- linki,
- ...

Warianty to dokumenty różniące się między sobą w niewielki sposób i będące zwykle związane jakąś relacją semantyczną.

Przykłady:

- nowelizacje aktów prawnych,
- dokumentacja kolejnych wersji urzędnienia.

Idea: zapobieganie redundancji części dokumentu wspólnych dla wszystkich wariantów.

Warianty treści: przykład

Do 31 grudnia 2011:

```
...  
<artykuł nr="212">  
  Przywóz wody nie jest  
  objęty cłem.</artykuł>  
...
```

Po 1 stycznia 2012:

```
...  
<artykuł nr="212">  
  Przywóz wody jest  
  objęty cłem  
  20%.</artykuł>  
...
```



Dokument wynikowy:

```
...  
<wariant id="a3819"/>  
...
```

Dokument wynikowy:

```
...  
<artykuł nr="212"  
  do="2011-12-31">  
  Przywóz wody nie jest  
  objęty cłem.  
</artykuł>  
<artykuł nr="212"  
  od="2012-01-01">  
  Przywóz wody jest  
  objęty cłem  
  20%.</artykuł>
```

Czy system zarządzania dokumentami powinien być jednocześnie systemem publikacyjnym (zawierać moduł publikacyjny)?

- + przecież publikacja jest celem!
- + po co przechowywać informację zwrotną z systemu składu (np. informacji o łamaniu stron) poza CMS-em?
- repozytorium jest niezależne, skład powinien odbywać się w wyspecjalizowanym systemie,
- procesy zarządzania dokumentami nie powinny być podporządkowane kształtowi przyszłej publikacji.

Zarządzanie:

- strukturą witryny (serwisami, kolekcjami stron, odnośnikami do innych stron),
- zawartością tekstową witryny (artykułami, stronami),
- wyglądem witryny (szablony stron).

Specyfika:

- dostępność w postaci aplikacji webowej,
- zawartość: zwykle fragmenty (X)HTML-a/XML-a i obrazki,
- łatwość tworzenia i publikacji treści,
- proces obsługi publikacji,
- funkcja podglądu nieopublikowanych treści,
- serwowanie dynamiczne lub eksport do struktur statycznych.

Obsługa kancelarii – jednostki organizacyjnej odpowiedzialnej za przyjmowanie, dystrybucję wewnętrzną i wysyłkę dokumentów.

Specyfika:

- proces zgodny z instrukcją kancelaryjną – szczegółowym opisem procedur obowiązujących w danej instytucji,
- klasyfikacja dokumentów zgodna z jednolitym rzeczowym wykazem akt (JRWA),
- wsparcie informatyczne w dwóch wariantach:
 - tradycyjny obieg dokumentów papierowych, podczas gdy system przechowuje metadane dokumentów (metryczki):
 - identyfikacja za pomocą kodów kreskowych, RFID, ...
 - w systemie zapisane jest miejsce przechowywania dokumentów papierowych (regał/półka),
 - obieg dokumentów elektronicznych:
 - wytwarzane dokumenty powstają w formie elektronicznej,
 - dokumenty papierowe są skanowane (i ew. OCR-owane), a następnie zapisywane w systemie.

Specyfika:

- proces zgodny z instrukcją archiwizacyjną – szczegółowym opisem procedur archiwizacji dokumentów,
- rejestracja dokumentów na podstawie spisów zdawczo-odbiorczych,
- klasyfikacja dokumentów zgodna z jednolitym rzeczowym wykazem akt (JRWA),
- kategorie archiwalne:
 - A – dokument o trwałej wartości archiwalnej, przeznaczony do przekazania do archiwum państwowego,
 - B n – dokument mający czasowe znaczenie praktyczne, przechowywany w archiwum przez n lat (np. B50 – 50 lat),
 - BE n – dokument podlega ekspertyzie po upływie n lat.
- kontrolowane udostępnianie zasobów archiwalnych,
- brakowanie = kontrolowane usuwanie dokumentów bez wartości archiwalnej.

Potrzeby klienta:

System do zarządzania zasobami archiwalnymi od roku 1952 do chwili obecnej.

Archiwum przed wdrożeniem:

- 3 km dokumentów papierowych,
- archiwum zdjęć,
- materiały audio/video (setki godzin nagrań).

Rozwiązanie:

- system szyty na miarę (na bazie istniejących komponentów),
- specyficzny interfejs.

Cechy systemu:

- pełna obsługa procesów archiwalnych (rejestracja → zarządzanie → udostępnianie → ekspertyza → brakowanie),
- dostęp z poziomu przeglądarki,
- możliwość wymiany danych z systemami zewnętrznymi (obieg dokumentów).

Moduły:

- repozytorium dokumentów i metadanych,
- wirtualna biblioteka,
- proste wyszukiwanie (względem metadanych),
- zaawansowane raportowanie.

W zasadzie dwa projekty:


- **2003-2004:** implementacja i wdrożenie systemu,
- **2005-2009:** skanowanie i masowy import 7 mln stron dokumentów.

Architektura:

- Windows 2003 Server (wcześniej: Solaris),
- PostgreSQL (wcześniej: Oracle),
- JBoss,
- macierz dyskowa 12 TB (wcześniej: jukebox DVD, 100 slotów).

Archiwum Prezydenta RP - Windows Internet Explorer

Ark Edycja Wydruk Ułubione Narzędzia Pomoc

 **Archiwum Prezydenta RP**
oparte na technologii eXms

PRZYJMIOWANIE UODSTĘPIANIE BRAKOWANIE EKSPERTYZA PRZEKAZYWANIE INDEKSY RAPORTY ADMINISTRACJA

WPROWADZANIE I WERYFIKACJA PRZEDPOŁA EWIDENCJA

> Szczegóły spisu zdawczo-odbiorczego <

Szczegóły spisu zdawczo-odbiorczego
„Przejdź wyżej...”

Numery spisu
Numer spisu: 176
Zespół: Kancelaria Prezydenta RP
Jednostka organizacyjna: PVL
Daty skrajne: 1996-1997
Data przyjęcia akt: 2005-07-12
Data sporządzenia spisu: 1997-09-15
Wybrakowano:
Przekazano do AAN:
Uwagi:
Zapisz zmiany

Porządek spisu zdawczo-odbiorczego

Lp.	Znak teckki	Tytuł	Daty skrajne	Kategoria archiwalna	Liczba teczek (kaset, fotografii)	Położenie (regal/półka)	Ilość m.b.	Uwagi	Status
1	PVL_040	Zaproszenia	1997	A	1	/	0.0		
2	PVL_045	Kongresy, zjazdy jubileuszowe, sympozja organizowane przez różne instytucje i organizacje	1996-1997	A	1	/	0.0		
3	PVL_060	Podania i skargi	1997	A	1	/	0.0		
4	PVL_070	Współpraca z organami administracji państwowej	1997	A	1	/	0.0		
5	PVL_079	Współpraca z innymi instytucjami	1997	A	1	/	0.0		
6	PVL_001	Współpraca z organizacjami międzynarodowymi	1997	A	1	/	0.0		
7	PVL_031	Współpraca z Ministerstwem Spraw Zagranicznych	1997	A	1	/	0.0		
8	PVL_032	Ambasadorowie i inni przedstawiciele RP	1997	A	1	/	0.0		
9	PVL_037	Wyjazdy zagraniczne pracowników Kancelarii Prezydenta RP	1997	A	1	/	0.0		
10	PVL_055	Przyjazdy, delegacje, inne	1997	A	1	/	0.0		

zalogowany jako: rag ostatnio zalogowany: 2005-10-08 18:33:18

Przed rokiem 2005:

- system zarządzający metadanymi ustaw i projektów ustaw,
- dokumenty przechowywane w systemie plików w postaci plików MS Worda,
- ręczne wersjonowanie dokumentów,
- ręczne ujednolicanie tekstów.

Wyodrębnienie części procesowej i redakcyjnej:

- ① moduł serwera obsługujący procesy legislacyjne:
 - rejestracja nowych tekstów,
 - kontrola nad stanem prac,
 - obsługa wariantów projektów nowelizujących,
 - uruchamianie akcji edycji/podglądu/ujednolicania,
 - ...

- ② środowisko edycyjne (XMetaL):
 - zapis kompletu informacji w tekstach:
 - adresy publikacji,
 - daty: uchwalenia, wejścia w życie, obowiązywania zestawów przepisów,
 - powiązania nowelizacyjne (przywołania),
 - znaczniki definicji,
 - dostosowania edytora XML-owego (CSS/makra/skróty klawiaturowe),
 - integracja z repozytorium z użyciem protokołu WebDAV.

USTAWA

z dnia 30 czerwca 2000 r.

Prawo własności przemysłowej

Tytuł I

Przepisy ogólne

Art. 1.

1. Ustawa normuje:

- 1) stosunki w zakresie wynalazków, wzorów użytkowych, wzorów przemysłowych, znaków towarowych, oznaczeń geograficznych i topografii układów scalonych;
- 2) zasady, na jakich przedsiębiorcy mogą przyjmować projekty racjonalizatorskie i wynagradzać ich twórców;
- 3) zadania i organizację Urzędu Patentowego Rzeczypospolitej Polskiej, zwanego dalej „Urzędem Patentowym”.

2. Przepisy ustawy nie uchybiają ochronie przedmiotów, o których mowa w ust. 1 pkt 1, przewidzianej w innych ustawach.

Art. 2.

Zwalczanie nieuczciwej konkurencji reguluje odrębna ustawa.

Art. 3.

1. Ilekroć w ustawie jest mowa o:

- 1) osobie – rozumie się przez to osobę fizyczną lub prawną;
- 2) osobie zagranicznej – rozumie się przez to osobę niemającą obywatelstwa polskiego i odpowiednio miejsca zamieszkania albo siedziby bądź poważnego przedsiębiorstwa na obszarze Rzeczypospolitej Polskiej;

USTAWA
z dnia 7 grudnia 2006 r.

**o zmianie ustawy o ujawnianiu informacji o dokumentach organów
bezpieczeństwa państwa z lat 1944-1990 oraz treści tych dokumentów**

Art. 1.

W ustawie z dnia 18 października 2006 r. o ujawnianiu informacji o dokumentach organów bezpieczeństwa państwa z lat 1944-1990 oraz treści tych dokumentów (Dz.U. Nr 218, poz. 1592) wprowadza się następujące zmiany:

1) art. 1 otrzymuje brzmienie:

„Art. 1. Ustawa określa zasady i tryb ujawniania informacji o dokumentach organów bezpieczeństwa państwa oraz treści tych dokumentów, znajdujących się w archiwum Instytutu Pamięci Narodowej – Komisji Ścigania Zbrodni przeciwko Narodowi Polskiemu, zwanego dalej „Instytutem Pamięci Narodowej” oraz tryb składania oraz oceny zgodności z prawdą oświadczeń dotyczących pracy lub służby w organach bezpieczeństwa państwa lub współpracy z tymi organami w okresie od dnia 22 lipca 1944 r. do dnia 31 lipca 1990 r.”;

2) po art. 3 dodaje się art. 3a w brzmieniu:

„Art. 3a. 1. Współpracą w rozumieniu ustawy jest świadoma i tajna współpraca z ogniwami operacyjnymi lub śledczymi organów bezpieczeństwa państwa w charakterze tajnego informatora lub pomocnika przy operacyjnym zdobywaniu informacji.

2. Współpracą w rozumieniu ustawy jest również działanie, którego obowiązek wynikał z ustawy obowiązującej w czasie tego działania w związku z pełnioną funkcją, zajmowanym stanowiskiem, wykonywaną pracą lub pełnioną służbą, jeżeli informacje przekazywane organom bezpieczeństwa państwa stwarzały zagrożenie dla wolności i praw człowieka i

The screenshot shows the XMetaL Author application window. The title bar reads "XMetaL Author - [http://localhost:8080/ksrp-webdav/Workspace/Ref_eda1d8310a0002f00b09321e219e30a.xml]". The menu bar includes "plik", "Edycja", "Widok", "Wstaw", "Narzędzia", "Okno", and "Pomoc". The toolbar contains various icons for file operations, editing, and navigation. The main editing area displays the following text with XML tags:

USTAWA
z dnia 29 sierpnia 2003 r.

o zmianie ustawy o Lasach Państwowych

Art. 1.

W ustawie z dnia 26 marca 1982 r. o Lasach Państwowych wprowadza się następujące zmiany:

1) w art. 2 w ust. 1 wyrazy „lub za popełnienie przestępstwa” „zastępuje się wyrazami „za popełnienie przestępstwa lub przestępstwa skarbowego”

2) w art. 9 w ust. 9 wyrazy „uznaje się za nieprzedstawiony przez Komisję Odpowiedzialności Konstytucyjnej” zastępuje się wyrazami „nie podlega rozpatrzeniu”;

3) w art. 10:

- a) w ust. 1a po wyrazie „podjęcia” dodaje się wyrazy „przez Zgromadzenie Narodowe”;
- b) w ust. 2 wyrazy „odpowiada warunkom wymaganym” zastępuje się wyrazami „spełniać warunki wymagane przez ustawę”;

4) w art. 11:

- a) w ust. 3 w zdaniu pierwszym po wyrazie „uchwałę” dodaje się wyraz „Sejmu”;
- b) w ust. 4 po wyrazie „uchwały” dodaje się wyraz „Sejmu”;

5) w art. 13:

- a) dodaje się ust. 2c w brzmieniu: „2c. Jeżeli Komisja Odpowiedzialności Konstytucyjnej występuje z wnioskiem o umorzenie w stosunku do Prezydenta postępowania z powodu przedawnienia lub śmierci, wówczas przewodniczący Zgromadzeniu Narodowemu stwierdza umorzenie postępowania w sprawie.”;
- b) dodaje się ust. 4 w brzmieniu: „4. Jeżeli Komisja Odpowiedzialności Konstytucyjnej występuje z wnioskiem o umorzenie w stosunku do osób, o których mowa w art. 1 ust. 1 pkt 2-7 albo ust. 2, postępowania z powodu przedawnienia lub śmierci osoby, której postępowanie dotyczy, wówczas Marszałek Sejmu stwierdza umorzenie postępowania w sprawie.”;

Art. 2.

Ustawa wchodzi w życie po upływie 14 dni od dnia ogłoszenia.

The status bar at the bottom shows "Rules Checking On" and "NUM".

Założenia ogólne:

- brak konstrukcji formatujących,
- reprezentacja struktury:
 - hierarchia jednostek: księga, dział, rozdział, artykuł, ...
 - zawartość tekstowa poszczególnych części aktu,
 - tytuły części tytułowanych,
 - przypisy końcowe,
- elementy dodatkowe:
 - uwagi do tekstu,
 - oznaczenia definicji,
 - dodatkowe oznaczenia orzeczeń TK,
 - elementy zewnętrzne w postaci wysp edycyjnych,
- znaczniki nowelizacyjne.

```
<ustawa data-uchwalenia="20 grudnia 1990">
  <tytul>o jednostkach innowacyjno-wdrożeniowych</tytul>
  <adres-publicacji>
    <adres-dzu rok="1991" numer="2" pozycja="7"/>
  </adres-publicacji>
  <artykul nr="1">Urząd Postępu Naukowo-Technicznego
    i Wdrożeń na dotychczasowych zasadach:
    <punkt nr="1">dokonuje skreśleń w rejestrze jednostek
      innowacyjno-wdrożeniowych w okresie trzech lat
      od dnia wejścia w życie niniejszej ustawy,</punkt>
    <punkt nr="2">wpisuje jednostki do rejestru.</punkt>
  </artykul>
  <artykul nr="2">Ustawa wchodzi w życie po upływie
    14 dni od dnia ogłoszenia.</artykul>
</ustawa>
```

Obecnie:

- dokumenty tworzone w XML-u (zgodnie z tzw. „schematem sejmowym”),
- prezentacja tekstów w wielu formatach (PDF, HTML, DOCX) obsługiwana za pomocą XSLT,
- wersjonowanie przy każdym zapisie; teksty przechowywane w bazie danych (Oracle, CLOB),
- mechanizm automatycznego ujednolicania (przywołania, znaczniki nowelizacyjne, ścieżki nowelizacyjne),
- konwerter Word2XML – prawie automatyczny,
- architektura: Solaris (Windows)/Oracle/JBoss/XMetaL.

```
ustawa:before  
{ content: "USTAWA\Az dnia " attr(data-uchwalenia) " r.";  
  font-weight: bold;  
  display: block;  
  text-align: center; }
```

```
<xsl:template match="artykul">
  <xsl:choose>
    <xsl:when test="count(ancestor::*[name() = 'dodaj'
      or name() = 'zastap']) > 0">
      <div class="artykul-cytowany">
        <xsl:if test="position() = 1">„</xsl:if>
        Art. <xsl:value-of select = "@nr"/>
        <xsl:apply-templates/>
      </div>
    </xsl:when>
    <xsl:otherwise>
      <div class="artykul">
        <div class="naglowek-artykulu">Art.
          <xsl:value-of select = "@nr"/></div>
          <xsl:apply-templates/>
        </div>
      </xsl:otherwise>
    </xsl:choose>
  </xsl:template>
```

Projekty nowelizacji można nakładać na teksty obowiązujące:

- linki nowelizacyjne pomiędzy parami tekstów zapisane w treści nowelizacji:
 - OO: przywołania = linki między ustawą nowelizującą a nowelizowaną,
 - CC: znaczniki nowelizacyjne reprezentujące pojedyncze zmiany wraz z datami ich obowiązywania;
 - cele linków zapisane w postaci wyrażeń XPath,
- możliwość nałożenia wielu projektów nowelizacji na pojedynczy tekst obowiązujący,
- wykrywanie potencjalnych konfliktów nowelizacyjnych.

Przywołania przechowują informację o linku między fragmentem tekstu nowelizacji a ustawą bazową. Są wstawiane w miejscu odwołania do tytułu ustawy nowelizowanej i zastępują go; informacja o ustawie pobierana jest przy każdym otwarciu tekstu bezpośrednio z bazy danych.

Przykład:

```
<przywołanie ustawa-id="8ab282971830bb8a01183221c82c0010"
      typ="nowelizacyjne">
  <ustawa-info data-uchwalenia="6 grudnia 1996">
    <tytul>o zastawie rejestrowym
      i rejestrze zastawów</tytul>
    <adres-publicacji>
      <adres-dzu rok="1996" numer="149" pozycja="703"/>
    </adres-publicacji>
  </ustawa-info>
  <przypis nr="2"/>
</przywołanie>
```

Dwa wymiary znaczników:

1 rodzaj operacji:

- <dodaj>,
- <zmien>,
- <usun>,

2 cel operacji:

- element struktury,
- fragment tekstu.

Tekst:

W art. 7 w ust. 1 dodaje się pkt 6a w brzmieniu:
„6a) analizę wykorzystania wynikających z kontroli wniosków dotyczących stanowienia lub stosowania prawa.”.

Reprezentacja XML-owa:

W art. 7 w ust. 1 dodaje się pkt 6a w brzmieniu:

```
<dodaj>  
  <po-elemencie id="//artykul[@nr='7']  
    /ustep[@nr='1']/punkt[@nr='6']"/>  
  <element>  
    <punkt nr="6a">analizę wykorzystania  
      wynikających z kontroli wniosków dotyczących  
      stanowienia lub stosowania prawa.</punkt>  
  </element>  
</dodaj>
```

Tekst:

W art. 2 ust. 3 wyrazy „armie” zastępuje się wyrazami „siły zbrojne” .

Reprezentacja XML-owa:

W art. 2 ust. 3 wyrazy

```
<zastap>
```

```
  <w-elemencie id="//artykul[@nr='2']  
                /ustep[@nr='3']"/>
```

```
  <tekst>armie</tekst>
```

```
  <akcja>zastępuje się wyrazami</akcja>
```

```
  <tekstem>siły zbrojne</tekstem>
```

```
</zastap>.
```

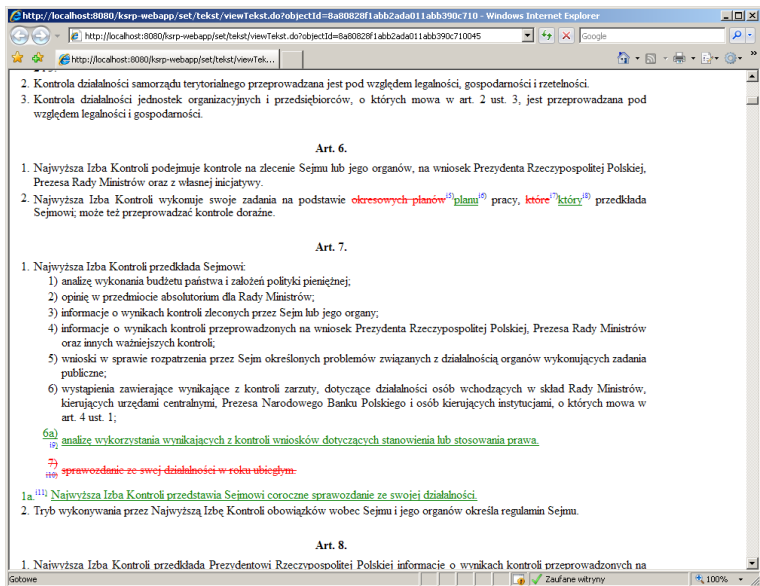
Model znacznika <zastap>

```
<xsd:element name="zastap">
  <xsd:complexType>
    <xsd:choice>
      <xsd:sequence>
        <xsd:choice maxOccurs="unbounded">
          <xsd:element ref="element"/>
          <xsd:element ref="elementy"/>
        </xsd:choice>
        <xsd:element ref="elementem" minOccurs="0"/>
      </xsd:sequence>
      <xsd:sequence>
        <xsd:element ref="w-elemencie"
          minOccurs="0" maxOccurs="unbounded"/>
        <xsd:element ref="tekst"/>
        <xsd:element ref="akcja" minOccurs="0"/>
        <xsd:element ref="tekstem"/>
      </xsd:sequence>
    </xsd:choice>
    <xsd:attributeGroup ref="atr-nowelizacyjne"/>
  </xsd:complexType>
</xsd:element>
```

5 modeli zawartości – do reprezentacji:

- 1 zmiany brzmienia pojedynczego elementu struktury:
 - `<element id="ścieżka">` – element pusty, ze ścieżką do elementu, którego zawartość ma zostać wymieniona,
 - `<elementem>` – nowe brzmienie elementu struktury (zapisane w treści znacznika),
- 2 zmiany brzmienia ciągłego zestawu elementów struktury:
 - `<elementy od="ścieżka1" do="ścieżka2">` – ze ścieżkami wskazującymi początek i koniec ciągłego zestawu elementów o zmienianym brzmieniu,
 - `<elementem>` – nowe brzmieniem zestawu,
- 3 zmiany brzmienia tekstu:
 - `<w-elemencie id="ścieżka">` – ze ścieżką do elementu, w którym ma zostać zmieniony tekst,
 - `<tekst>` – tekst, który ma zostać zastąpiony nowym,
 - `<akcja>` – tekst łączący (np. „zastępuje się wyrazy”),
 - `<tekstem>` – nowe brzmienie zastępowanych wyrazów,

- 4 obudowania tekstu elementem struktury:
 - `<element id="ścieżka">` – ze ścieżką do elementu, którego tekst ma zostać obudowany elementem struktury,
 - `<elementem poziom="element">` – ze wskazaniem nazwy elementu reprezentującego poziomu struktury, którego należy użyć do obudowania tekstu (np. *ustęp*),
- 5 obniżenia tekstu do poziomu tekstu podstawowego:
 - `<element id="ścieżka">` – ze ścieżką do elementu, którego zawartość ma zostać obniżona.



2. Kontrola działalności samorządu terytorialnego przeprowadzana jest pod względem legalności, gospodarności i rzetelności.

3. Kontrola działalności jednostek organizacyjnych i przedsiębiorców, o których mowa w art. 2 ust. 3, jest przeprowadzana pod względem legalności i gospodarności.

Art. 6.

1. Najwyższa Izba Kontroli podejmuje kontrole na zlecenie Sejmu lub jego organów, na wniosek Prezydenta Rzeczypospolitej Polskiej, Prezesa Rady Ministrów oraz z własnej inicjatywy.

2. Najwyższa Izba Kontroli wykonuje swoje zadania na podstawie **okresowych planów⁽¹⁾ planu⁽²⁾ pracy, które⁽³⁾ który⁽⁴⁾** przedkłada Sejmowi; może też przeprowadzać kontrole doraźne.

Art. 7.

1. Najwyższa Izba Kontroli przedkłada Sejmowi:

- 1) analizę wykonania budżetu państwa i założeń polityki pieniężnej;
- 2) opinię w przedmiocie absoltorium dla Rady Ministrów;
- 3) informacje o wynikach kontroli zleconych przez Sejm lub jego organy;
- 4) informacje o wynikach kontroli przeprowadzonych na wniosek Prezydenta Rzeczypospolitej Polskiej, Prezesa Rady Ministrów oraz innych ważniejszych kontroli;
- 5) wnioski w sprawie rozpatrzenia przez Sejm określonych problemów związanych z działalnością organów wykonujących zadania publiczne;
- 6) wystąpienia zawierające wynikające z kontroli zarzuty, dotyczące działalności osób wchodzących w skład Rady Ministrów, kierujących urzędami centralnymi, Prezesa Narodowego Banku Polskiego i osób kierujących instytucjami, o których mowa w art. 4 ust. 1;

^(6a) ⁽⁷⁾ [analizę wykorzystania wynikających z kontroli wniosków dotyczących stanowienia lub stosowania prawa.](#)

⁽⁷⁾ ~~sprawozdanie ze swej działalności w roku ubiegłym.~~

^(1a) ⁽¹¹⁾ [Najwyższa Izba Kontroli przedstawia Sejmowi coroczne sprawozdanie ze swojej działalności.](#)

2. Tryb wykonywania przez Najwyższą Izbę Kontroli obowiązków wobec Sejmu i jego organów określa regulamin Sejmu.

Art. 8.

1. Najwyższa Izba Kontroli przedkłada Prezydentowi Rzeczypospolitej Polskiej informacje o wynikach kontroli przeprowadzonych na

Gotowe Zaufane wtrymy 100%

Większość tekstów jest wciąż tworzonych w formatach MS Worda. Muszą zostać skonwertowane do formatu XML-owego:

- pliki Worda zapisywane w formacie DOCX,
- proces konwersji zaimplementowany w Javie,
- oprócz struktury weryfikowane dodatkowe własności tekstu (np. interpunkcja),
- błędy konwersji prezentowane w postaci komentarzy Worda,
- dokument wynikowy tworzony na bieżąco,
- wyrażenia regularne dostępne jako parametr – mogą być modyfikowane przez klienta.

USTAWA z dnia 3 lipca 2002 r.

Prawo lotnicze

Rozdział 1 Zakres regulacji i definicje

1. Przepisy Prawa lotniczego regulują stosunki prawne z zakresu lotnictwa cywilnego.)
2. Przepisy Prawa lotniczego stosuje się do polskiego lotnictwa cywilnego oraz, w zakresie ustalonym przez to prawo, również do obcego lotnictwa cywilnego.
3. Lotnictwo cywilne obejmuje wszystkie rodzaje lotnictwa, z wyjątkiem lotnictwa państwowego, to jest państwowych statków powietrznych oraz lotnisk państwowych wykorzystywanych wyłącznie do startów i lądowań państwowych statków powietrznych.)
4. Przepisów prawa lotniczego nie stosuje się do lotnictwa państwowego, z wyjątkiem przepisów art. 1 ust. 6-7²⁾, art. 2, 3-10, 11 ust. 1, art. 12, 14, 33, 35 ust. 2, art. 43, 44, 60, 74, 75, art. 76 ust. 2, art. 89, 92, 119-125, 128, 130, 133, 140, 149 i 150 ustawy, oraz z zastrzeżeniem ust. 5,
5. W przypadku wykorzystywania:
- 1) polskich i obcych państwowych statków powietrznych do prowadzenia działalności innej niż służba publiczna;
 - 2) lotnisk lotnictwa państwowego do startów i lądowań cywilnych statków powietrznych oraz statków, o których mowa w pkt 1
- przepisy ustawy stosuje się odpowiednio.

[6. (uchylony).]

Komentarz [D1]: Znaczenie elementu 'statki', s spodziewano się jednego z następujących elementów: '(oddział, średyna, artykuł, uwaga, wyspa)'.

Komentarz [D2]: Element ma za duzo wcięcia

Komentarz [D3]: Niektórzy mają zdania elementu zmuszającego? Okazanie średnika.

Czyli czego nie ma, a co można:

- prezentacja każdego tekstu w wersji „na dany dzień” ,
- wyszukiwanie pełnotekstowe (i ew. bardziej zaawansowane),
- obsługa prac komisji,
- ...