

Systemy zarządzania treścią

Szymon Ziolo
szioło@mimuw.edu.pl

Statystyka

• **90%** zasobów informacyjnych firm
jest przechowywanych w dokumentach
a nie w bazach danych (Deloitte & Touche)

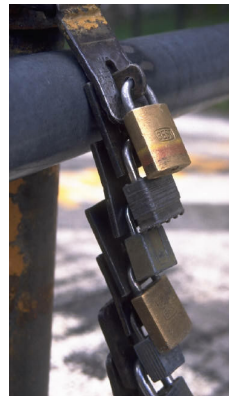
• **92 miliardy** dokumentów
tworzonych co roku (AIIM)

Najczęściej stosowane „systemy” zarządzania dokumentami

- Tradycyjny system obiegu dokumentów papierowych (szafy, segregatory, asystentka, gонец).
- Poczta elektroniczna, wymiana przy pomocy dyskietek, pen-drive’ów, itp.
- Współdzielony system plików (dysk sieciowy).

Dostępność

- Większość z istniejących dokumentów, mimo iż dostępnych elektronicznie, jest nadal nieużyteczna.
- Powody:
 - zamknięte i niekompatybilne formaty (n.p. DTP, Ms Word),
 - niedostępne/nieznane miejsce przechowywania,
 - rozproszone, rozbieżne i nieaktualne wersje,
 - nieużywane już aplikacje.



Kiedy przestaje wystarczać system plików

- Zasoby informacyjne:
 - o dużej objętości,
 - o skomplikowanej strukturze i powiązaniach,
 - o dużej wartości,
 - o długim cyklu życia informacji,
 - o dużej częstotliwości aktualizacji informacji.
- Organizacja:
 - wieloosobowe zespoły,
 - wysoka specjalizacja członków zespołu,
 - rozproszenie geograficzne.

Kiedy przestaje wystarczać system plików

- Przykłady:
 - wydawnictwo encyklopedyczne,
 - wydawnictwo prawnicze,
 - wydawca czasopism,
 - koncern przemysłowy, producent zaawansowanych technicznie urządzeń,
 - operator rozległej sieci telekomunikacyjnej, energetycznej, ...,
 - organizacja oparta na wiedzy,
 - administracja państwowa.

Proste (i tanie) rozwiązania

- Centralne repozytoria, np.:
 - CVS (Concurrent Versions System),
 - SVN (Subversion).
- 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.
- Rozwiązania typu wiki, np. MediaWiki, MoinMoin.

Budowa typowego systemu zarządzania treścią

- Repozytorium dokumentów.
- Warstwa aplikacji:
 - funkcjonalność biznesowa,
 - system przepływu prac,
 - silnik wyszukiwania,
 - silnik sprawdzający poprawność dokumentów,
 - silniki przekształceń,
 - system publikacyjny.
- Interfejs użytkownika:
 - system nawigacji,
 - system edycyjny.

Repozytorium

- Przechowywanie dokumentów:
 - dowolne typy dokumentów,
 - wersjonowanie,
 - blokowanie dokumentów do edycji,
 - specjalne wsparcie dla SGML-a/XML-a.
- Metainformacje:
 - informacje o dokumentach (np. autorzy, data publikacji, wersja),
 - przechowywane poza dokumentami (np. w bazie relacyjnej),
 - konfigurowalne w dowolny sposób,
 - metainformacje strukturalne (listy, struktury),
 - synchronizacja metainformacji z zawartością dokumentu,
 - wyszukiwanie w metainformacjach.

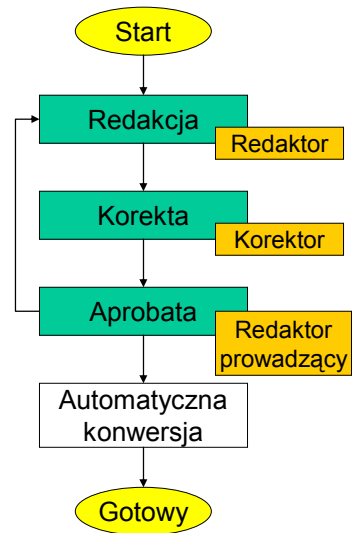
Metainformacje – przykład

```
<zeznanie-sprawcy>
Wypadek dnia <data>
13.01.2001r</data>
o godzinie <godzina>13.13
</godzina> (<dzien-tygodnia>
piątek</dzien-tygodnia>) miał
miejsce nie z mojej winy.
<poszkodowany>Alojzy
M.</poszkodowany> nie miał
żadnego pomysłu w którą
stronę uciekać, więc go
przejechałem.
</zeznanie-sprawcy>
```

Nr protokołu: 1313/2001
Miejsce: Dołowice Górne
Data: 13.01.2001
Czas: 13.13
Rodzaj: potrącenie pieszego
Sprawca: Walenty Pechowy
Sprawę prowadzi: st. asp. Jan Łapówka

Przeływ prac

- Dwa podejścia:
 - tradycyjnie: działania osób „popychają” przepływ prac.
 - elektroniczny przepływ prac: steruje działaniami osób.



Interfejs użytkownika

- Nawigacja w zasobach repozytorium:
 - drzewiasta struktura katalogów,
 - wyświetlanie (wybranych) metainformacji,
 - struktura katalogów może przenosić istotną informację semantyczną o zależnościach między obiektami.
- Edycja dokumentów:
 - SGML/XML: edytor strukturalny,
 - obiekty dowolnego typu: aplikacje przeznaczone do ich edycji.

System uniwersalny

- Konfiguracja:
 - typów dokumentów:
 - SGML/XML: każdy schemat (DTD) definiuje osobny typ,
 - katalogów i ich dopuszczalnej zawartości,
 - metainformacji,
 - typów dowiązań,
 - schematu przepływu prac,
 - uprawnień,
 - ...

System otwarty

- Możliwość implementacji logiki biznesowej na bazie funkcjonalności systemu:
 - operacje wyzwalane przed lub po standardowych operacjach,
 - przedefiniowanie standardowych operacji,
 - dodawanie nowych funkcjonalności.
- Wymagania:
 - API pozwalające na dostęp do obiektów logiki systemu,
 - rozszerzalny interfejs użytkownika.

Wdrożenie systemu zarządzania dokumentami

- Analiza wymagań:
 - konfrontacja wymagań z podstawową funkcjonalnością systemu,
 - wybór systemu zarządzania dokumentami.
- Projektowanie:
 - projekt konfiguracji (w tym: typy dokumentów, DTD),
 - projekt warstwy logiki biznesowej,
 - projekt implementacji logiki biznesowej w oparciu o logikę systemu.
- Implementacja logiki biznesowej.
- Wdrożenie:
 - instalacja, konfiguracja,
 - integracja z innymi systemami (np. systemem składu),
 - konwersja zastanych danych, zasilenie systemu.
- Szkolenia.
- Pielęgnacja, wsparcie.

Rodzaje i odmiany systemów zarządzania dokumentami/treścią

- *Web Content Management Systems* – zarządzanie zawartością witryny internetowej.
- *Enterprise Content Management Systems* – zarządzanie dokumentami biznesowymi organizacji.
- System obiegu dokumentów kancelaryjnych.
- System publikacyjny.
- Portal korporacyjny.
- System do pracy grupowej.
- Elektroniczne archiwum.

Dwa podejścia do zarządzania dokumentami

- Podejście treściowo-centriczne – zarządzanie treścią:
 - wszystkie zasoby dostępne dla (uprawnionych) użytkowników,
 - użytkownik decyduje, z których zasobów w danej chwili korzysta,
 - typowy sposób dostępu:
 - przeglądanie katalogów,
 - wyszukiwanie.
- Podejście procesowe – obieg dokumentów:
 - ściśle zdefiniowane role i kompetencje,
 - użytkownik wykonuje zadania wskazane przez system,
 - system zarządza przekazywaniem dokumentu (zadania) kolejnym osobom,
 - typowy sposób dostępu: lista zadań do wykonania.

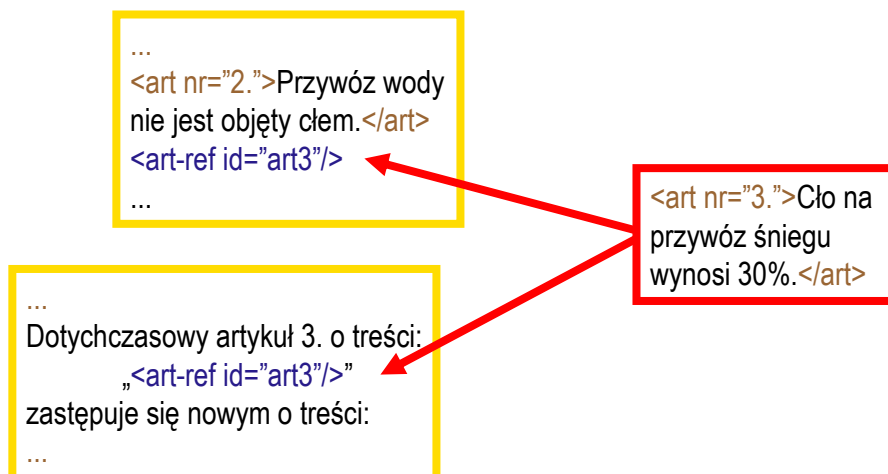
Dwa podejścia do zarządzania dokumentami



Zarządzanie treścią w wydawnictwie

- Treść – podstawowy produkt biznesowy wydawnictwa.
- Typowe problemy:
 - planowanie zawartości publikacji i zarządzanie przygotowaniem treści,
 - wielokrotne wykorzystanie zasobów treści:
 - w różnych publikacjach,
 - w różnych mediach;
 - personalizacja dostarczanej treści.
- Ciekawe przypadki:
 - wydawnictwo encyklopedyczne, słownikowe,
 - wydawca zbiorów aktów prawnych,
 - wydawca czasopism.

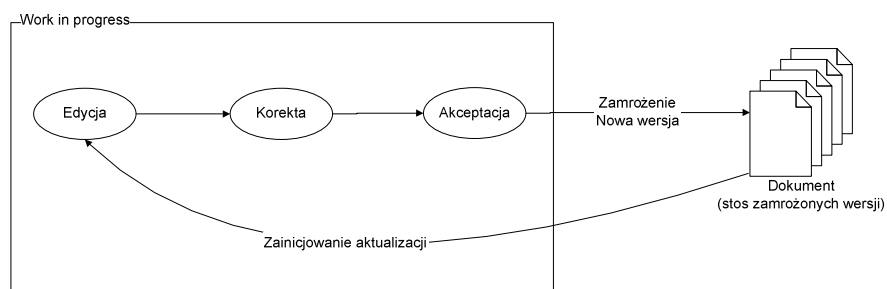
Subdokumenty



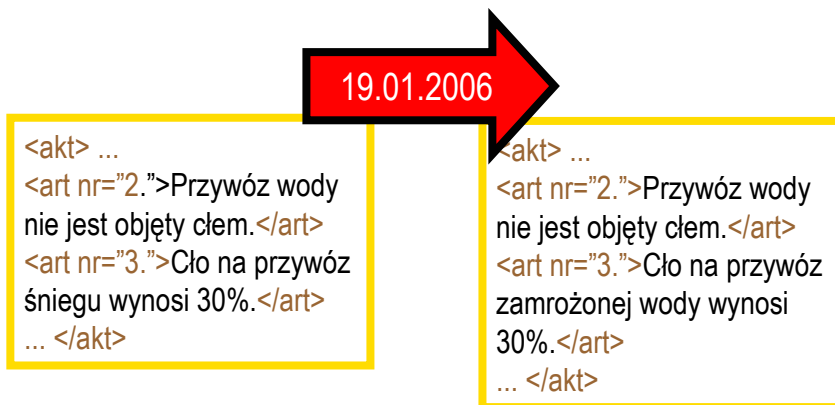
Zarządzanie wersjami

- Przechowywanie wersji dokumentu z kolejnych edycji:
 - możliwość łatwego powrotu do wcześniejszej wersji.
- Wydania (*releases*) / rewizje:
 - dokument w momencie publikacji jest zamrażany,
 - aby wprowadzić zmiany, trzeba założyć nowe wydanie,
 - nowe wydanie przechodzi cały cykl życia w systemie, aż do publikacji.
- Warianty:
 - zarządzanie zmianami semantycznymi w dokumentach,
 - przykłady:
 - nowelizacje aktów prawnych,
 - dokumentacja kolejnych wersji urzędzenia;
 - zapobieganie redundancji części dokumentu wspólnych dla wszystkich wersji.

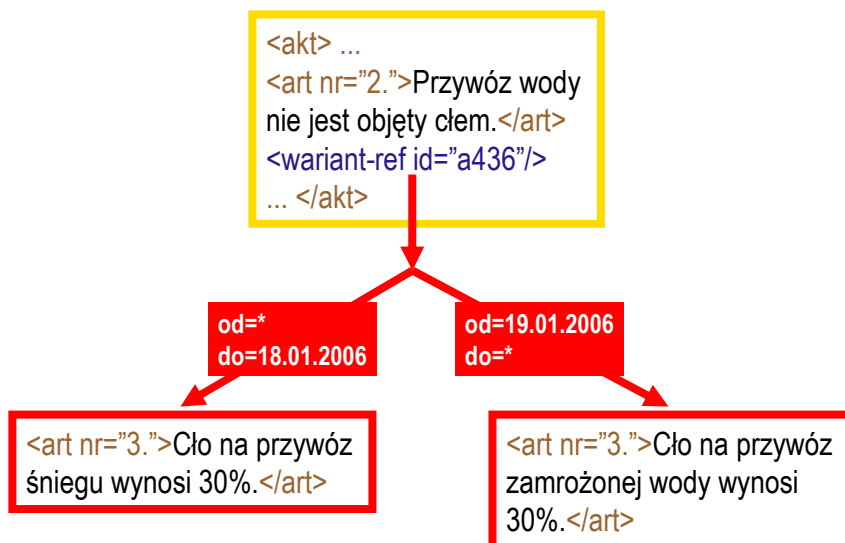
Wydania / rewizje



Warianty dokumentów (1)



Warianty dokumentów (2)



Aktualizacja „twardych faktów”

- „Twarde” fakty:
 - dane liczbowe, statystyczne,
 - okresowo się zmieniające lub uaktualniane.
- Ta sama informacja (np. liczba mieszkańców Danii) występuje w wielu miejscach:
 - w tekście hasła „Dania”,
 - w tabeli pokazującej zmiany liczby mieszkańców Danii w czasie,
 - w zbiorczej tabeli danych krajów europejskich.
- Jak nie dopuścić do redundancji:
 - przechowujemy „twarde” fakty w jednej kopii w tabeli,
 - modelujemy w dokumentach miejsce ich wstawienia.

Aktualizacja „twardych faktów”

```
<haslo id="dk"><tytul>Dania</tytul>
...
<treść>Liczba mieszkańców:
<wstaw-akt-lm/> (<wstaw-akt-lm-rok/>).</treść>
</haslo>
```

```
select akt_lm from liczba_mieszk
where panstwo="dk"
```

```
select akt_lm_rok from liczba_mieszk
where panstwo="dk"
```

Zarządzanie dokumentami vs. publikowanie

- Czy system zarządzania dokumentami powinien być jednocześnie systemem publikacyjnym (zawierać system publikacyjny)?
- Za:
 - zarządzamy po to, aby opublikować,
 - niekiedy konieczność przechowania informacji zwrotnej z systemu składu (np. informacji o łamaniu stron).
- Przeciw:
 - istnieją wyspecjalizowane systemy składu,
 - niebezpieczeństwo podporządkowania procesów zarządzania dokumentami kształtowi przyszłej publikacji.