



20 listopada 2003


Systemy zarządzania treścią




Statystyka

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

92 miliardy dokumentów
tworzonych co roku (ASM)



Technologia



Miejsce:

- rylca,
- dluta,
- pióra,
- maszyny do pisania

zajął komputer.

Ale czy w istocie wiele się zmieniło?

- kalka → *copy*
- gumka → *back-space*

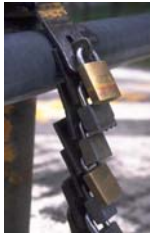




Dostępność

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, ...
- administracja państwowa.



Budowa systemu zarządzania dokumentami

Repozytorium dokumentów.

Warstwa aplikacji:

- system przepływu prac,
- silnik wyszukiwania,
- silnik walidujący dokumenty,
- silniki przekształceń,
- system publikacyjny.

Interfejs użytkownika:

- system nawigacji,
- system edycyjny.



Użytkownicy i bezpieczeństwo

Autentykacja i autoryzacja użytkowników.

Grupy użytkowników.

Wspólna praca zespołu użytkowników:

- pobieranie dokumentów do edycji (check-out),
- zwracanie zmodyfikowanych dokumentów (check-in).

Uprawnienia:

- do wykonania poszczególnych operacji,
- do nawigacji/przeglądania informacji różnego rodzaju:
 - dokumentów, drzew katalogów,
 - metainformacji,
 - linków.
 - ...
- zależność od stanu przepływu prac.



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 (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>
```

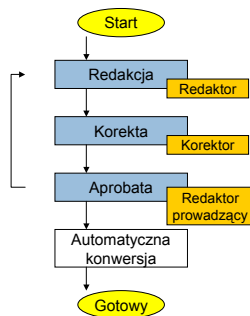
Nr protokołu: 1313/2001
Miejsce: Dołowice Górne
Data: 13.01.2001
Czas: 13.13
Rodzaj: potrącenie pieszego
Sprawca: Walenty Pechowy
Sprawę prowadzi:
st. asp. Jan Łąpówka



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.



Przykład: sigmalink

Projekt publikacyjny

Obiekty informacyjne

Title	Date
Whitepaper, English	March 1997
Getting Started	30 Jul 1995
Report	

Katalog

Metainformacje

empolis

Przydatne cechy XML-a

Oddzielenie **znaczenia tekstu** od sposobu jego prezentacji.

Zalety:

- jedno źródło,
- wiele różnych publikacji,
- redukcja nadmiarowości,
- uproszczone zarządzanie i aktualizacja.

Przydatne cechy XML-a

Stworzenie **najodpowiedniejszego modelu** dla naszych informacji.

Wyciskamy maksimum informacji z dokumentów:

- wyszukiwanie,
- kategoryzacja,
- automatyczne przetwarzanie:
 - indeksy,
 - linkowanie,
 - nowe dokumenty na podstawie starych.



empolis

Wyszukiwanie

Wyszukiwanie w metainformacjach.

Wyszukiwanie pełnotekstowe – z możliwością ograniczenia obszaru wyszukiwania do wybranych elementów:

- Słowacki:
 - Słowacki Raj
 - słowacki polityk
 - ...
- Słowacki in nazwisko:
 - Słowacki, Juliusz



empolis

Wyszukiwanie – inny pomysł

Wykorzystanie modelu wiedzy o przeszukiwanej rzeczywistości.

Szukaj: niebieski opel corsa

80% podobierstwa

Sprzedam <model> opia corsa z 1999 roku. Tel. (22) 123 45 67.

kolor niebieski jest podobny do granatowego w 80%

Sprzedam <kolor>granatowego </kolor> <marka>opla</marka> <model>corse</model> z <rocznik>1999</rocznik> roku. Tel. (22) 123 45 67.

empolis

Dowiązania

Rodzaje dowiązań:

- między (całymi) dokumentami,
- między zawartościami dokumentów,
- jedno- lub dwukierunkowe,
- o dwóch lub więcej końcach.

Funkcje dowiązań:

- dołączanie grafiki,
- dołączanie subdokumentów,
- zarządzanie wersjami (wariantami),
- linki semantyczne, przenoszące znaczenie specyficzne dla projektu:
 - typ linku,
 - metainformacje na linkach.

empolis

Linki semantyczne

Powiązanie dokumentów (fragmentów dokumentów) z innymi dokumentami (fragmentami).

Nadanie znaczenia linkom:

- wyszukiwanie z użyciem faktów,
- efektywna nawigacja,
- automatyczne przetwarzanie.

```
<OSOBA-MOWIACA><KSIAZE ID-KRAJU="d1">Hamlet</KSIAZE></OSOBA-MOWIACA>
<WYPOWIEDZ><KRAJ ID="d1">Dania</KRAJ> jest więzieniem. </WYPOWIEDZ>
książę - kraj
```



Linki semantyczne – przykłady

Encyklopedia:

- urodzony w,
- syn – ojciec,
- tworzył pod wpływem,
- jest/był prezydentem.

Prawo:

- na podstawie aktu,
- zmiana aktu,
- uchylene aktu.

Dokumentacja techniczna:

- składa się z,
- nadklasa – podklasa.



Subdokumenty

```
...
<art nr="2.">Przywóz wody
nie jest objęty cłem.</art>
<art-ref id="art3"/>
...
```

```
<art nr="3.">Cło na
przywóz śniegu
wynosi 30%.</art>
```

```
...
Dotychczasowy artykuł 3. o treści:
„<art-ref id="art3"/>”
zastępuje się nowym o treści:
...
```



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.

Zarządzanie zmianami semantycznymi w dokumentach:

- przykłady:
 - nowelizacje aktów prawnych,
 - dokumentacja kolejnych wersji urzędzenia;
- zapobieganie redundancji części dokumentu wspólnych dla wszystkich wersji.



Warianty dokumentów (1)

```
<akt> ...
<art nr="2.">Przywóz wody
nie jest objęty cłem.</art>
<art nr="3.">Cło na przywóz
śniegu wynosi 30%.</art>
... </akt>
```

20.11.2003

```
<akt> ...
<art nr="2.">Przywóz wody
nie jest objęty cłem.</art>
<art nr="3.">Cło na przywóz
zamrożonej wody wynosi
30%.</art>
... </akt>
```



Warianty dokumentów (2)

```
<akt> ...
<art nr="2.">Przywóz wody
nie jest objęty cłem.</art>
<wariant-ref id="a436"/>
... </akt>
```

od="+
do="19.11.2003

```
<art nr="3.">Cło na przywóz
śniegu wynosi 30%.</art>
```

od="20.11.2003
do="+

```
<art nr="3.">Cło na przywóz
zamrożonej wody wynosi
30%.</art>
```



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" – przykład

```
<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"
```



Operacje na dokumentach i ich zawartości

Split / join:

- podział dokumentu SGML/XML na kilka dokumentów zgodnie ze strukturą dokumentu,
- niezależny byt dokumentów składowych,
- połączenie w całość.

Wirtualne dokumenty:

- kolekcja dokumentów określonych typów,
- traktowana jako pojedynczy dokument:
 - check-out: złożenie treści wirtualnego dokumentu,
 - check-in: rozbicie zawartości na dokumenty składowe.



Import i eksport

Import:

- interaktywny: możliwość zaimportowania pojedynczych dokumentów przez użytkowników,
- masowy:
 - zasilanie systemu zastanymi danymi,
 - konwersje do SGML-a / XML-a.

Eksport:

- publikowanie zawartości.



System generyczny

Konfiguracja:

- typów dokumentów:
 - SGML/XML: każda 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,
- przededefiniowanie 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.

