Wykład 2

Systemy zarządzania dokumentami i treścią

Systemy Zarządzania Treścią
Politechnika Poznańska

Część 1

System uniwersalny i otwarty

System uniwersalny

System uniwersalny zarządzania dokumentami charakteryzuje się szeroką konfigurowalnością — to, jakie treści może obsługiwać, nie jest z góry narzucone, lecz definiowane przez użytkownika.

Zakres konfiguracji

Element konfiguracji	Opis
Typy dokumentów	W SGML/XML każdy schemat (DTD) definiuje osobny typ dokumentu
Katalogi	Definicja katalogów i ich dopuszczalnej zawartości
Metainformacje	Definicja struktury metadanych opisujących dokumenty
Typy dowiązań	Określenie rodzajów powiązań między dokumentami
Schemat przepływu prac	Definicja procesów pracy (workflow)
Uprawnienia	Definicja praw dostępu dla użytkowników i grup

> *Uniwersalność oznacza, że ten sam system może obsługiwać akty prawne, hasła encyklopedyczne czy dokumentację techniczną — wystarczy inna konfiguracja. Definiowanie typów dokumentów (DTD, XML Schema) omówione w **wykładach 3–8**.*

System otwarty

System otwarty umożliwia **implementację logiki biznesowej** na bazie funkcjonalności systemu. Użytkownik nie jest ograniczony do standardowych operacji.

Możliwości rozszerzania

Możliwość	Opis
Operacje wyzwalane	Operacje wykonywane przed lub po standardowych operacjach systemu
Przedefiniowanie operacji	Nadpisywanie standardowych operacji własną logiką
Dodawanie funkcjonalności	Rozszerzanie systemu o nowe możliwości

Wymagania otwartości

1. **API** pozwalające na dostęp do obiektów logiki systemu,
2. **Rozszerzalny interfejs użytkownika**.

> *Otwartość systemu jest kluczowa w dużych wdrożeniach, gdzie wymagania biznesowe wykraczają poza standardową funkcjonalność.*

Wdrożenie systemu zarządzania dokumentami

Wdrożenie systemu zarządzania dokumentami przebiega w kilku etapach:

1. Analiza wymagań

- konfrontacja wymagań z podstawową funkcjonalnością systemu,
- wybór systemu zarządzania dokumentami.

2. Projektowanie

- projekt konfiguracji (w tym: typy dokumentów, DTD),
- projekt warstwy logiki biznesowej,
- projekt implementacji logiki biznesowej w oparciu o logikę systemu.

3. Implementacja

- implementacja logiki biznesowej.

4. Wdrożenie

- instalacja, konfiguracja,
- integracja z innymi systemami (np. systemem składu),
- konwersja zastanych danych, zasilenie systemu.

5. Utrzymanie

- szkolenia,
- pielęgnacja, wsparcie.

Rodzaje i odmiany systemów zarządzania dokumentami/treścią

Systemy zarządzania treścią występują w wielu odmianach, dostosowanych do różnych zastosowań:

Rodzaj systemu	Przeznaczenie
Web Content Management Systems	Zarządzanie zawartością witryny internetowej
Enterprise Content Management Systems	Zarządzanie dokumentami biznesowymi organizacji
System obiegu dokumentów kancelaryjnych	Obsługa korespondencji i spraw urzędowych
System publikacyjny	Wspomaganie procesu wydawniczego
Portal korporacyjny	Organizacja zasobów i usług informacyjnych firmy

Rodzaj systemu	Przeznaczenie
System do pracy grupowej	Współpraca zespołów nad wspólnymi dokumentami
Elektroniczne archiwum	Długoterminowe przechowywanie dokumentów

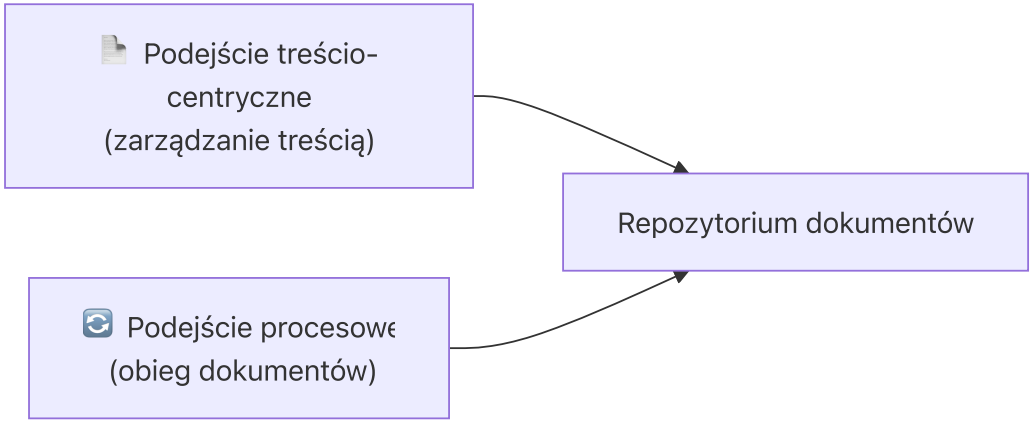
---

Część 2

Dwa podejścia do zarządzania dokumentami

Podejście treścio-centryczne vs procesowe

W zarządzaniu dokumentami wyróżnia się dwa zasadniczo odmienne podejścia:



Podejście treścio-centryczne

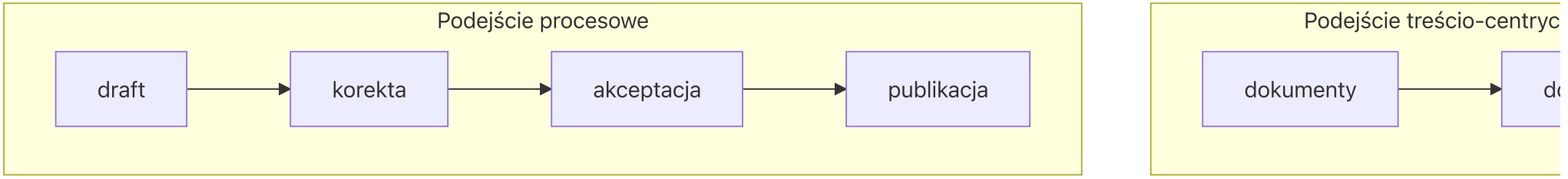
Cecha	Opis
Zasoby	Wszystkie zasoby dostępne dla (uprawnionych) użytkowników
Decyzja	Użytkownik decyduje, z których zasobów w danej chwili korzysta

Cecha	Opis
Dostęp	Przeglądanie katalogów, wyszukiwanie

Podejście procesowe

Cecha	Opis
Role	Ściśle zdefiniowane role i kompetencje
Zadania	Użytkownik wykonuje zadania wskazane przez system
Przepływ	System zarządza przekazywaniem dokumentu (zadania) kolejnym osobom
Dostęp	Lista zadań do wykonania

Dwa podejścia — ilustracja



Podsumowanie różnic

Kryterium	Treścio-centryczne	Procesowe
Inicjatywa	Użytkownik (pull)	System (push)
Struktura	Katalogi, wyszukiwanie	Kroki procesu
Role	Elastyczne	Ściśle zdefiniowane
Zastosowanie	Repozytoria, archiwa	Obieg dokumentów, akceptacje

> *W praktyce dojrzałe systemy łączą oba podejścia: repozytorium treści (treścio-centryczne) + mechanizm workflow (procesowe).*

Zarządzanie treścią w wydawnictwie

W wydawnictwie treść jest podstawowym produktem biznesowym. To specyficzny kontekst, który stawia szczególne wymagania systemom CMS.

Typowe problemy wydawnictwa

Problem	Opis
Planowanie zawartości	Planowanie publikacji i zarządzanie przygotowywaniem treści
Wielokrotne wykorzystanie	Ponowne użycie zasobów treści w różnych publikacjach i mediach
Personalizacja	Dostosowywanie dostarczanej treści do odbiorcy

Ciekawe przypadki

- Wydawnictwo encyklopedyczne, słownikowe — ta sama treść w wielu publikacjach,
- Wydawca zbiorów aktów prawnych — złożone zależności między dokumentami,
- Wydawca czasopism — cykliczny obieg treści.

Część 3

Dowiązania i wersjonowanie

Dowiązania

Dowiązania (linki) są podstawowym mechanizmem wyrażania relacji między dokumentami i ich fragmentami.

Rodzaje dowiązań

Rodzaj	Opis
Między dokumentami	Powiązania całych dokumentów
Między zawartościami	Powiązania fragmentów dokumentów
Jedno- lub dwukierunkowe	Kierunek nawigacji
O dwóch lub więcej końcach	Liczba powiązanych elementów

Funkcje dowiązań

- Dołączanie grafiki,
- Dołączanie subdokumentów,
- Zarządzanie wersjami (wariantami),
- Linki semantyczne — przenoszące znaczenie specyficzne dla projektu:
  - typ linku,
  - metainformacje na linkach.

Linki semantyczne

Linki semantyczne to powiązania dokumentów (lub ich fragmentów) z innymi dokumentami (fragmentami), które nadają znaczenie tym powiązaniom.

Zalety nadawania znaczenia linkom

- Wyszukiwanie z użyciem faktów — możliwość zadawania zapytań opartych na relacjach,
- Efektywna nawigacja — poruszanie się po sieci powiązań,
- Automatyczne przetwarzanie — maszynowe wykorzystanie relacji.

Przykład semantycznego dowiązania

```
<OSOBA-MOWIACA>
  <KSIAZE ID-KRAJU="d1">Hamlet<KSIAZE>
</OSOBA-MOWIACA>
```

```
<WYPOWIEDZ>
  <KRAJ ID="d1">Dania</KRAJ> jest więzieniem.
</WYPOWIEDZ>
```

W tym przykładzie atrybut `ID-KRAJU` ustanawia semantyczny związek między postacią a krajem — relację "książę → kraj".

Linki semantyczne — przykłady dziedzinowe

Różne dziedziny wymagają różnych typów linków semantycznych:

Dziedzina	Przykłady linków
Encyklopedia	urodzony w, syn–ojciec, tworzył pod wpływem, jest/był prezydentem
Prawo	na podstawie aktu, zmiana aktu, uchylenie aktu
Dokumentacja techniczna	składa się z, nadklasa–podklasa

> *Typ linku semantycznego jest zależny od ontologii dziedziny — to ona określa, jakie relacje są istotne i jakie niosą znaczenie.*

Subdokumenty

Subdokumenty umożliwiają wielokrotne wykorzystanie fragmentów dokumentów w różnych miejscach, bez kopiowania treści.

Problem redundancji

Bez mechanizmu subdokumentów ta sama treść musiałaby być kopiowana w wielu miejscach, co prowadzi do:

- niespójności danych,
- trudności w aktualizacji,
- nadmiarowości.

Rozwiązanie: odwołania

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

<!-- Odwołanie w innym miejscu -->
<art-ref id="art3" />
```

Subdokument jest definiowany raz i odwoływany wielokrotnie — zmiana w źródle automatycznie aktualizuje wszystkie miejsca użycia.

Zarządzanie wersjami

Zarządzanie wersjami dokumentu obejmuje trzy powiązane mechanizmy:

1. Wersje

- przechowywanie wersji dokumentu z kolejnych edycji,
- możliwość łatwego powrotu do wcześniejszej wersji.

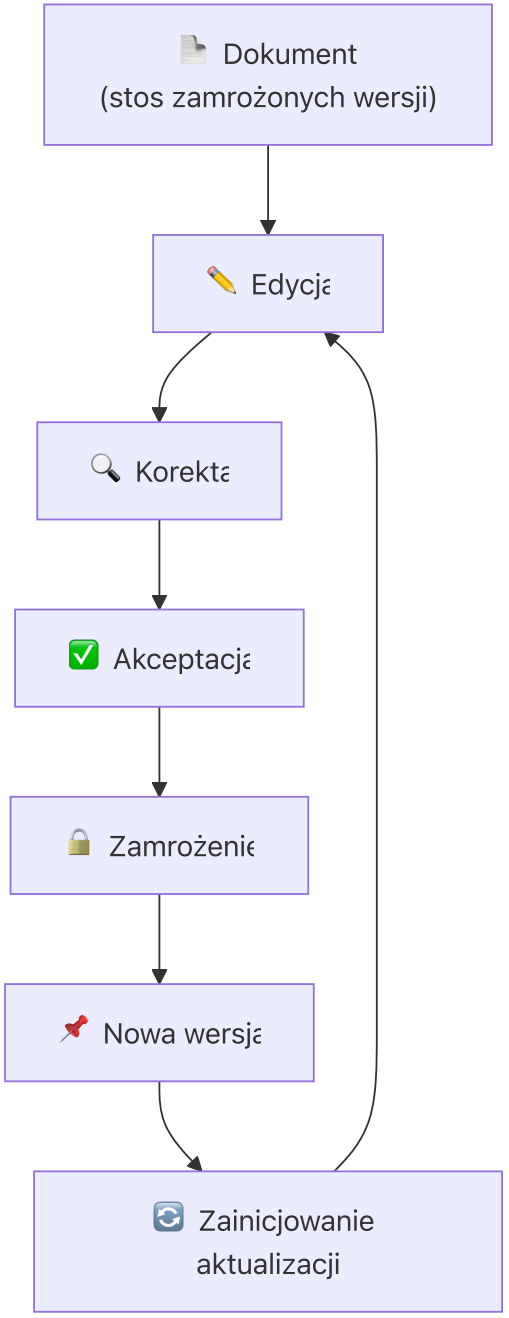
2. Wydania (releases) / rewizje

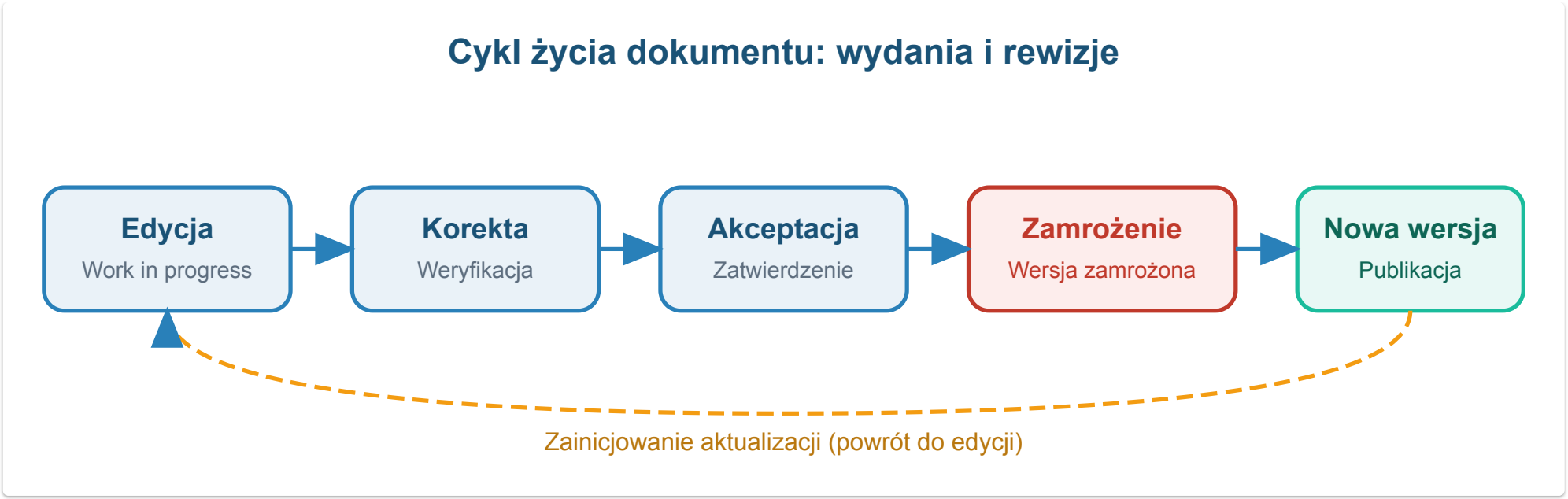
- dokument w momencie publikacji jest zamrażany,
- aby wprowadzić zmiany, trzeba założyć nowe wydanie,
- nowe wydanie przechodzi cały cykl życia w systemie, aż do publikacji.

3. Warianty

- zarządzanie zmianami semantycznymi w dokumentach,
- przykłady: nowelizacje aktów prawnych, dokumentacja kolejnych wersji urządzenia,
- zapobieganie redundancji części dokumentu wspólnych dla wszystkich wersji.

Wydania i rewizje — cykl życia


Dokument
(stos zamrożonych wersji)
Edycja
Korekta
Akceptacja
Zamrożenie
Nowa wersja
Zainicjowanie
aktualizacji



> *Kluczowa idea: po zamrożeniu dokument nie jest modyfikowany — każda zmiana tworzy nową wersję, co gwarantuje integralność historyczną.*

Warianty dokumentów

Warianty pozwalają reprezentować różne wersje treści obowiązujące w różnych okresach — bez duplikowania wspólnych części.

Przykład: nowelizacja aktu prawnego

```
<!-- Wariant obowiązujący od 19.01.2006 -->
<akt>
  <art nr="3.">Cło na przywóz zamrożonej wody wynosi 30%.</art>
</akt>

<!-- Warianty z odniesieniem czasowym -->
<art nr="3.">Cło na przywóz zamrożonej wody wynosi 30%.</art>  od=19.01.2006 do=*
<art nr="3.">Cło na przywóz śniegu wynosi 30%.</art>         od=* do=18.01.2006
```

Zasada

Wariant definiuje się raz, z zakresem obowiązywania (`od` – `do`). Wspólne części dokumentu nie są powielane — odwołują się do wspólnych zasobów.

CMS wielojęzyczny — warianty językowe

Mechanizm wariantów ma bezpośrednie zastosowanie w systemach wielojęzycznych — treść w różnych językach to szczególny przypadek wariantów.

Wielojęzyczność jako wymaganie CMS

Opis przedmiotu wymienia *explicite* budowę systemu wielojęzycznego. Jest to kluczowe dla:

- zastosowań B2B i B2C (obsługa rynków zagranicznych),
- portali instytucjonalnych (obowiązek publikacji w kilku językach),
- serwisów międzynarodowych (jedna treść, wiele języków).

Strategie architektury wielojęzycznej

Strategia	Przykład	Zalety	Wady
Podkatalogi	`example.com/pl/` , `example.com/en/`	Proste, wspólne SEO domeny	Jedna baza, jeden hosting
Poddomeny	`pl.example.com` , `en.example.com`	Niezależne SEO per język	Konfiguracja DNS
Osobne instancje	Oddzielny CMS per język	Pełna niezależność	Podwójne utrzymanie
Headless + i18n	API z metadanymi języka	Elastyczność	Złożoność developmentu

Model treści wielojęzycznej

W systemach zarządzania dokumentami treść wielojęzyczna modeluje się jako warianty językowe tego samego dokumentu:

```
<haslo id="dk">
  <tytul lang="pl">Dania</tytul>
  <tytul lang="en">Denmark</tytul>
```

```
  <tytul lang="de">Dänemark</tytul>
</haslo>
```

Wyzwania wielojęzyczności

Wyzwanie	Opis
Synchronizacja wersji	Zmiana w jednym języku musi znaleźć odbicie w innych
Workflow tłumaczeń	Odrębny proces akceptacji dla każdego języka
Kompletność	Które języki mają kompletną treść, a które nie
Metadane językowe	Oznaczenie języka, kierunek pisma, lokalizacja
SEO wielojęzyczne	Odrębna optymalizacja pod wyszukiwarki per język

> *Wielojęzyczność to **nie tylko tłumaczenie treści** — to pełne zarządzanie wariantami językowymi, ich synchronizacją i procesem akceptacji.*

Aktualizacja "twardych faktów"

"Twarde" fakty to dane liczbowe, statystyczne, które okresowo się zmieniają lub są uaktualniane.

Problem redundancji

Ta sama informacja (np. liczba mieszkańców Danii) występuje w wielu miejscach:

- w tekście hasła "Dania",
- w tabeli pokazującej zmiany liczby mieszkańców w czasie,
- w zbiorczej tabeli danych krajów europejskich.

Rozwiązanie

- przechowujemy "twarde" fakty w jednej kopii w tabeli,
- modelujemy w dokumentach miejsce ich wstawienia.

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

Wartość jest pobierana w czasie generowania publikacji zapytaniem do bazy danych (np. `select akt_lm from liczba_mieszk where panstwo="dk"`).

Część 4

Zarządzanie a publikowanie. Pula zasobów

Zarządzanie dokumentami vs publikowanie

Czy system zarządzania dokumentami powinien być jednocześnie systemem publikacyjnym?

Argumenty "za"

- zarządzamy po to, aby opublikować,
- niekiedy konieczność przechowania informacji zwrotnej z systemu składu (np. informacji o łamaniu stron).

Argumenty "przeciw"

- istnieją wyspecjalizowane systemy składu,
- niebezpieczeństwo podporządkowania procesów zarządzania dokumentami kształtowi przyszłej publikacji.

> *Kluczowa zasada: **oddzielenie zarządzania treścią od publikacji**. Treść powinna być neutralna względem medium, w którym zostanie opublikowana.*

Paradygmat neutralnej puli zasobów



Istota paradygmatu

Treść jest przechowywana w neutralnej puli zasobów — niezależnej od konkretnej publikacji. Z tej samej puli można budować różne publikacje:

- różne książki,
- różne media (druk, web, e-book),
- różne warianty językowe.

> *Neutralność oznacza, że treść nie jest "skrojona" pod jedną publikację — dzięki temu może być wielokrotnie wykorzystana.*

Case study: Planeta Actimedia

Planeta Actimedia — jedno z największych hiszpańskojęzycznych wydawnictw encyklopedycznych, publikujące na rynek hiszpański oraz Ameryki Łacińskiej.

Działalność

- wydaje wiele encyklopedii,
- przygotowuje kilka encyklopedii w tym samym czasie.

Cele wdrożenia systemu zarządzania treścią

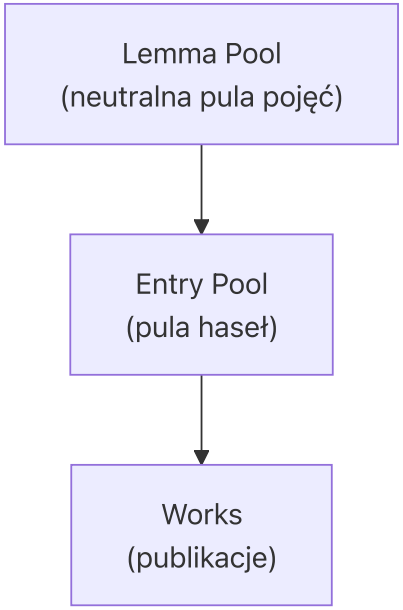
Cel	Opis
Kontrola struktury	Kontrola struktury treści haseł
Wielokrotne wykorzystanie	Ponowne użycie treści haseł w różnych encyklopediach
Wsparcie planowania	Planowanie zawartości encyklopedii

Cel	Opis
Automatyzacja współpracy	Zautomatyzowanie współpracy z zewnętrznymi autorami

---

Neutralna pula zasobów w Planecie

Planeta Actimedia stosuje wielowarstwową architekturę pul:



Trzy poziomy zasobów

Poziom	Zawartość	Przykład
Lemma Pool	Neutralne pojęcia (warianty)	"woda", "woda (chem.)", "woda (leks.)", "wódka"
Entry Pool	Hasła encyklopedyczne zbudowane z pojęć	"woda (encykl.)", "woda (słown.)"
Works	Konkretne publikacje (encyklopedie)	"Encyklopedia Powszechna", "Encyklopedia Chemii", "Słownik Jęz. Polskiego"

> *Pojęcie (np. "woda") istnieje raz w Lemma Pool i jest **automatycznie tworzone** jako hasło w różnych publikacjach.*

Organizacja pracy w Planecie

W Planecie działają **osobne zespoły**:

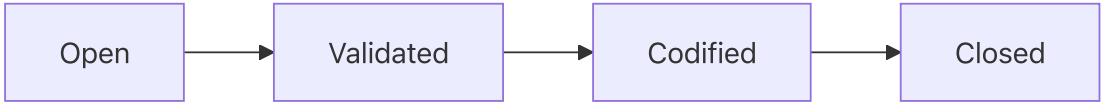
Zespół	Rola
Autorzy pojęć	Zewnętrzni współpracownicy i koordynatorzy tworzący pojęcia w Lemma Pool
Planerzy zawartości	Planowanie zawartości encyklopedii
Przygotowanie haseł	Tworzenie haseł encyklopedii w Entry Pool

Typowy proces pracy

1. Planer wybiera pojęcia, korzystając z informacji o ich objętości,
2. Planer może zlecić przygotowanie **wariantu pojęcia o zadanej objętości**,
3. System przypisuje (istniejące i nowe) warianty do encyklopedii,
4. Autorzy przygotowują treść brakujących wariantów pojęć,
5. System tworzy hasła złożone z wariantów pojęć,
6. Utworzone hasła są ręcznie dostosowywane.

Współpraca z zewnętrznymi autorami w Planecie

Cykl życia pojęcia w systemie odzwierciedla stan współpracy z autorem:



Stany cyklu życia

Stan	Opis
Open	Pojęcie otwarte do opracowania
Validated	Pojęcie zwalidowane
Codified	Pojęcie skodyfikowane (ujednolicone)
Closed	Pojęcie zamknięte

Każdy stan ma warianty wskazujące na status współpracy z autorem:

- Assigned — przypisane autorowi,
- Sent — wysłane do autora,
- Returned — zwrócone przez autora,
- Released — zwolnione (ukończone).

Case study: Wolters Kluwer Publication Build

Wolters Kluwer Publication Build — wspólny projekt czterech wydawnictw koncernu Wolters Kluwer.

Założenia

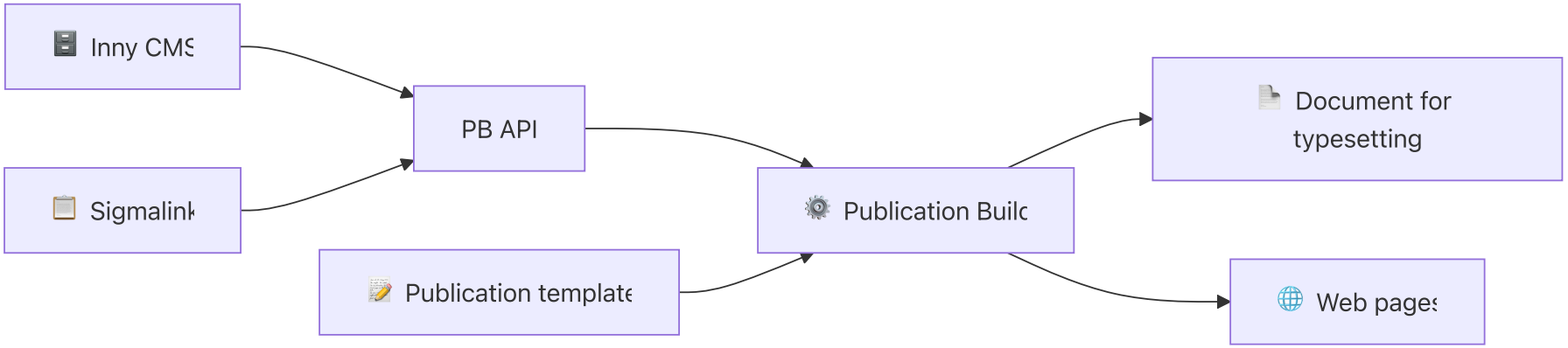
- system zarządzania dokumentami NIE zawiera systemu publikacyjnego,
- zarządzanie neutralną pulą zasobów,
- zasoby posiadają wartość dodaną (metainformacje, linki itp.).

Problem badawczy

- jak odwzorować strukturę zarządzanych zasobów na strukturę publikacji?
- jak zbudować wiele publikacji z tych samych zasobów?
- jak zrobić to w sposób uniwersalny?

> *To klasyczny przykład oddzielenia zarządzania treścią od publikacji w skali koncernu wydawniczego.*

Publication Build — schemat architektury



Komponenty

Komponent	Rola
Publication Build (PB)	Silnik budowania publikacji z neutralnych zasobów
PB API	Interfejs programistyczny do zasobów CMS
Publication template	Szablon definiujący strukturę publikacji
Web Content Service	Generowanie stron web
Rendition Service	Generowanie dokumentów do składu

Publication Build — przykład szablonu

Szablon publikacji definiuje, jak z neutralnych zasobów zbudować konkretną publikację:

```
<book>
  <title>Famous scientists</title>
  <cms:include-MIO id="145"/>
```

```
  <for-each-object query="get-objects-of-type('scientist')">
    <section id="this.id">
      <insert-content/>
      <p>This scientist was influenced by:</p>
      <for-each-object query="get-linked-objects(this,'influenced-by')">
        <reference id="this.id"/>
      </for-each-object>
      <for-each-object query="get-linked-objects(this,'author')">
        <subsection>
          <insert-content/>
        </subsection>
      </for-each-object>
    </section>
  </for-each-object>
</book>
```

> *Szablon wykorzystuje zapytania (`get-objects-of-type`, `get-linked-objects`) do pobierania zasobów z puli i budowania z nich publikacji.*

Podsumowanie

Temat	Kluczowe wnioski
System uniwersalny	Konfigurowalność typów dokumentów, katalogów, metadanych, dowiązań, workflow
System otwarty	API i rozszerzalny interfejs umożliwiają implementację logiki biznesowej
Dwa podejścia	Treścio-centryczne (pull) vs procesowe (push)
Dowiązania	Linki semantyczne przenoszą znaczenie specyficzne dla dziedziny
Wersjonowanie	Wersje, wydania (zamrożone), warianty (czasowe)
Twarde fakty	Jedna kopia w tabeli + miejsce wstawienia w dokumencie
Neutralna pula zasobów	Oddzielenie treści od publikacji, wielokrotne wykorzystanie
Case studies	Planeta Actimedia, Wolters Kluwer — separacja zarządzania od publikacji

Dziękuję za uwagę

Systemy Zarządzania Treścią

Politechnika Poznańska