

# Systemy Zarządzania Treścią

# Co to jest "zarządzanie treścią"?

- Komercyjny buzzword związany z ekspansją zastosowań Internetu (WWW) oraz rozwojem narzędzi służących do budowy aplikacji Internetowych.
- Istnieją w tej chwili dziesiątki (a może już i setki) systemów określanych jako "systemy zarządzania treścią" (Content Management Systems, CMS).
- Nie istnieje wyróżnik określający, co CMS ma zawierać. Poszczególne systemy różnią się zarówno oferowaną funkcjonalnością, jak i ceną (0\$ ↔ kilku mln.\$).
- Istnieje duży chaos w zakresie terminologii, standardów, technologii i biznesowej retoryki związanej z systemami CMS oraz ich zastosowaniami.

# Co to jest "treść"?

- Termin "treść" nie ma jednej definicji. Niektóre rodzaje:
  - Treść ukazująca się na ekranie przeglądarki: tekst, grafika, tło, dźwięk, animacja, video, przyciski, pola do wypełnienia, menu do wybrania, wykonywane na ekranie skrypty,...;
  - Odpowiedniki, elementy składowe lub generatory tej treści przechowywane po stronie serwera (np. w bazie danych);
  - Procesy, programy, reguły, metody, algorytmy pozwalające na generowanie treści z określonych źródeł, np. z bazy danych serwera lub z innych stron Web.
  - Różne formy opisu treści lub metadanych dotyczących treści, formaty, schematy, opisy dotyczące autorów treści, daty utworzenia, daty obowiązywania, własności, itd.
  - Różne formy kontroli i organizacji treści oraz usprawnienia dostępu: katalogi, klasyfikacje, indeksy, słowniki, ...

# Treść vs. dane, informacja, wiedza

- Jest dość trudno podać definicje różnicujące te terminy.
  - Są często używane jako synonimy.
  - Niekiedy odzwierciedlają emocjonalny stosunek do przedmiotu, np. dla ludzi AI "wiedza" lepiej pasuje do "inteligencji".
  - Niekiedy odzwierciedlają specyfikę celu przetwarzania i jakąś jego nową jakość: np: "wydobywanie wiedzy", a nie "wydobywanie danych"; "baza wiedzy", a nie "baza danych".
  - Niekiedy odzwierciedlają stosunek do procesów decyzyjnych zachodzących w ludzkim umyśle (dane muszą zamienić się w informację, ta zaś w wiedzę, aby mogła być podjęta decyzja).
- Treść jest rozumiana jako informacja, dane lub wiedza:
  - przekazywana do końcowego użytkownika przeglądarki,
  - zawarta w repozytorium serwera aplikacji internetowej.

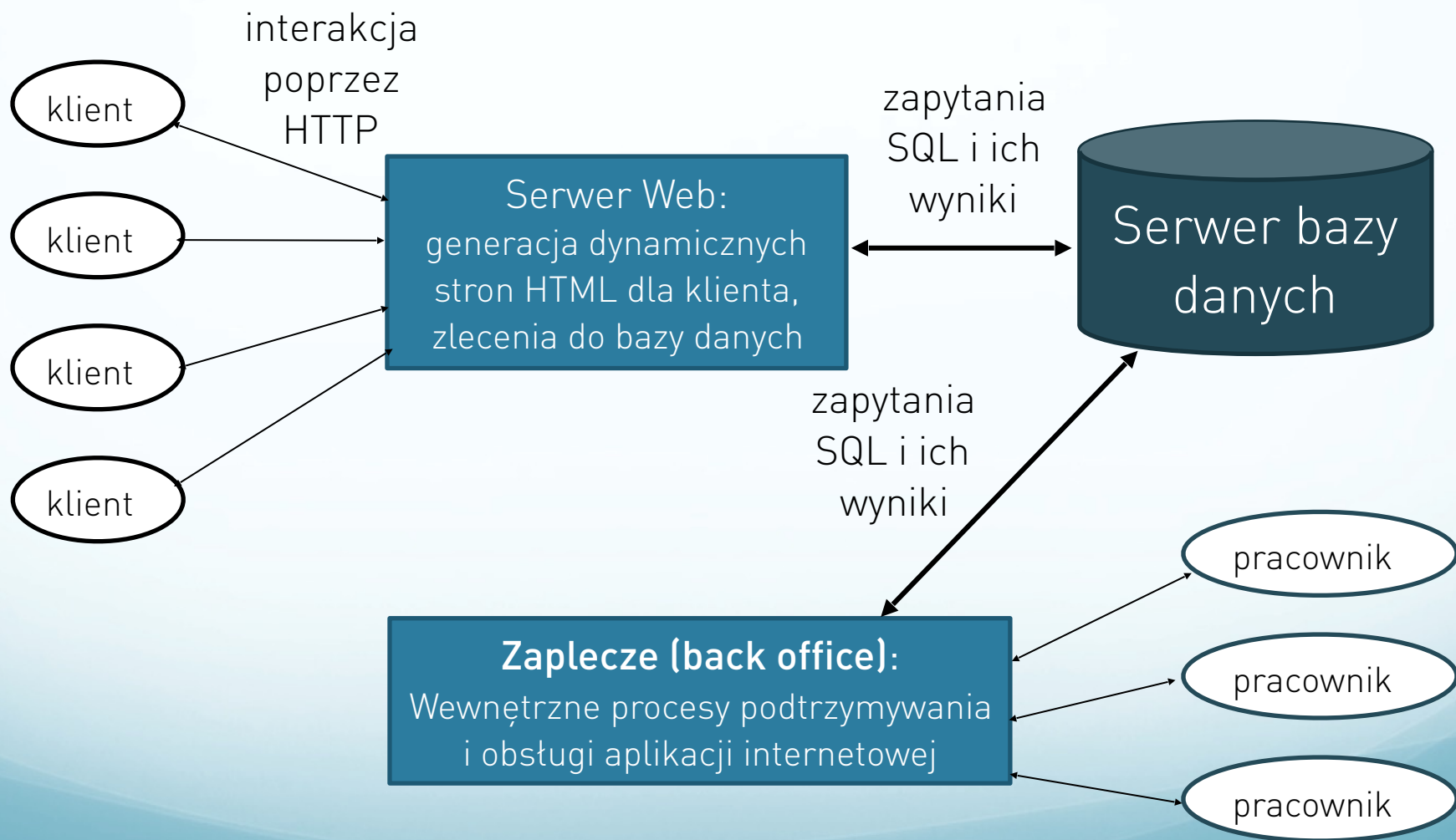
# Formaty i standardy treści

- **Setki formatów i standardów obowiązujących w zakresie reprezentacji, przechowywania, przetwarzania i udostępniania treści.**
  - Klasycznym standardem jest HTML, z licznymi rozszerzeniami w kierunku dynamizacji stron (JavaScript, aplety, ASP, JSP, ...)
  - Najnowszym buzzwordem jest XML oraz związane z nim technologie lub standardy (DTD, RDF, XSL, XSLT, XQL, SOAP, ...)
  - Wiele formatów reprezentacji tekstu: .txt, .doc, .rtf, .pdf, .ppt,...
  - Dziesiątki formatów graficznych (grafiki wektorowej i pikselowej), formatów animacji, audio i video.
  - Formaty, modele i standardy baz danych.
  - Formaty i standardy języków programowania stron internetowych (Java, SQL, ODBC, JDBC, PHP, Perl, Python, ...).

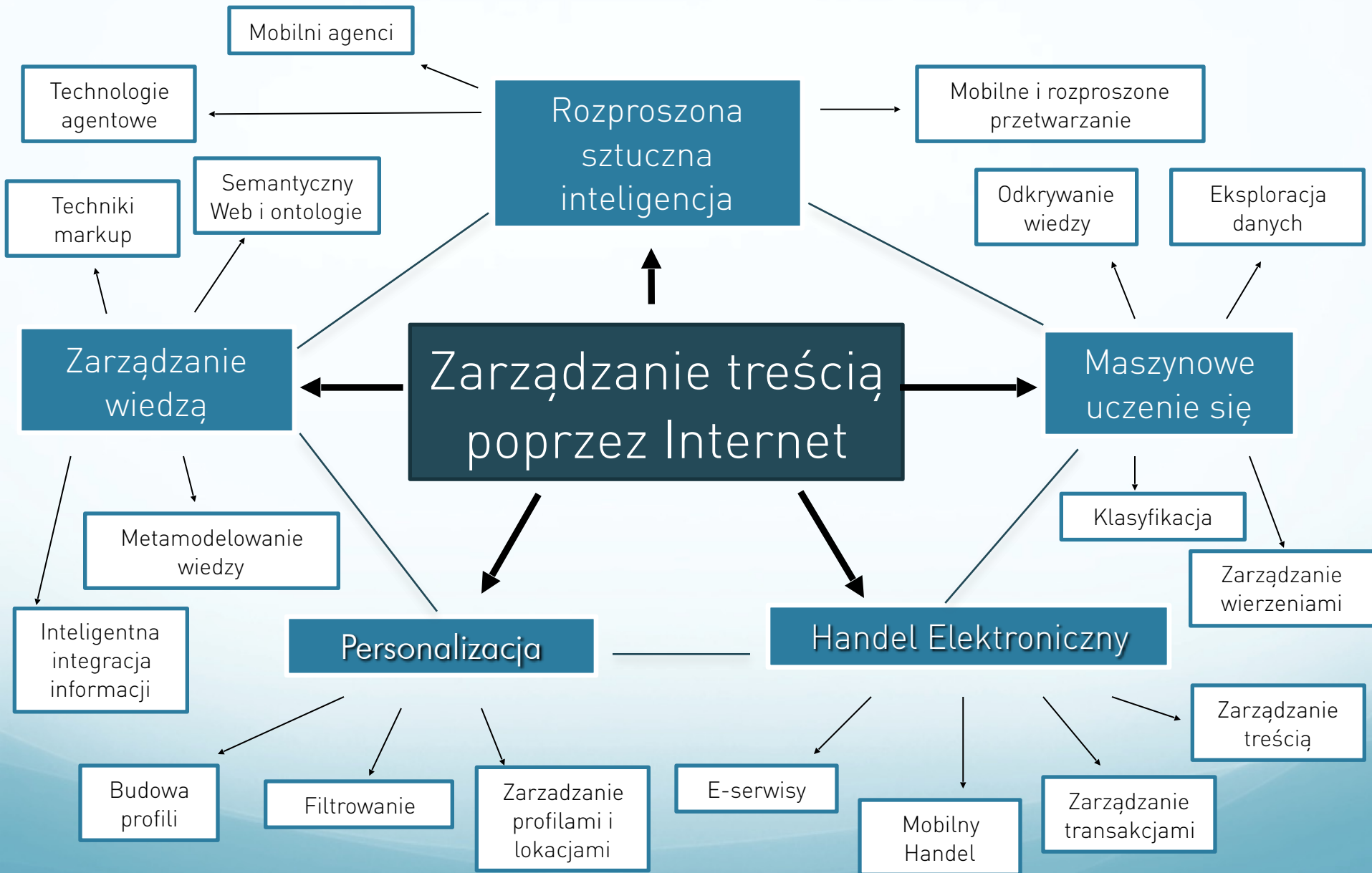
# Twórczy chaos w dziedzinie CMS

- Nowość technologiczna i rynkowa oraz możliwość zarobienia wielkich pieniędzy rodzi na początku chaos.
  - Jeżeli dla danego typu treści są popularne formaty A, B, C, to potrzebne będą odwzorowania  $A \rightarrow B$ ,  $B \rightarrow A$ ,  $A \rightarrow C$ , ... Liczba odwzorowań rośnie w kwadracie do liczby formatów.
  - Brak standardów i niekompatybilne rozwiązania implikuje oprogramowanie pośredniczące (middleware), bazujące najczęściej na nowym formacie (patrz CORBA, również XML).
  - Syndrom "dwóch programistów w garażu": sukces rozwiązania dla mikro-skali powoduje rozszerzanie bottom-up tego rozwiązania dla makro-skali (patrz HTML, XML, PHP, ...). Skutek: 1000-stronicowe podręczniki "prostego" języka XML.
- Twórczy chaos był już w innych dziedzinach (np. w językach programowania). Zwykle po pewnym czasie ustępuje.

# Techniczna architektura CMS

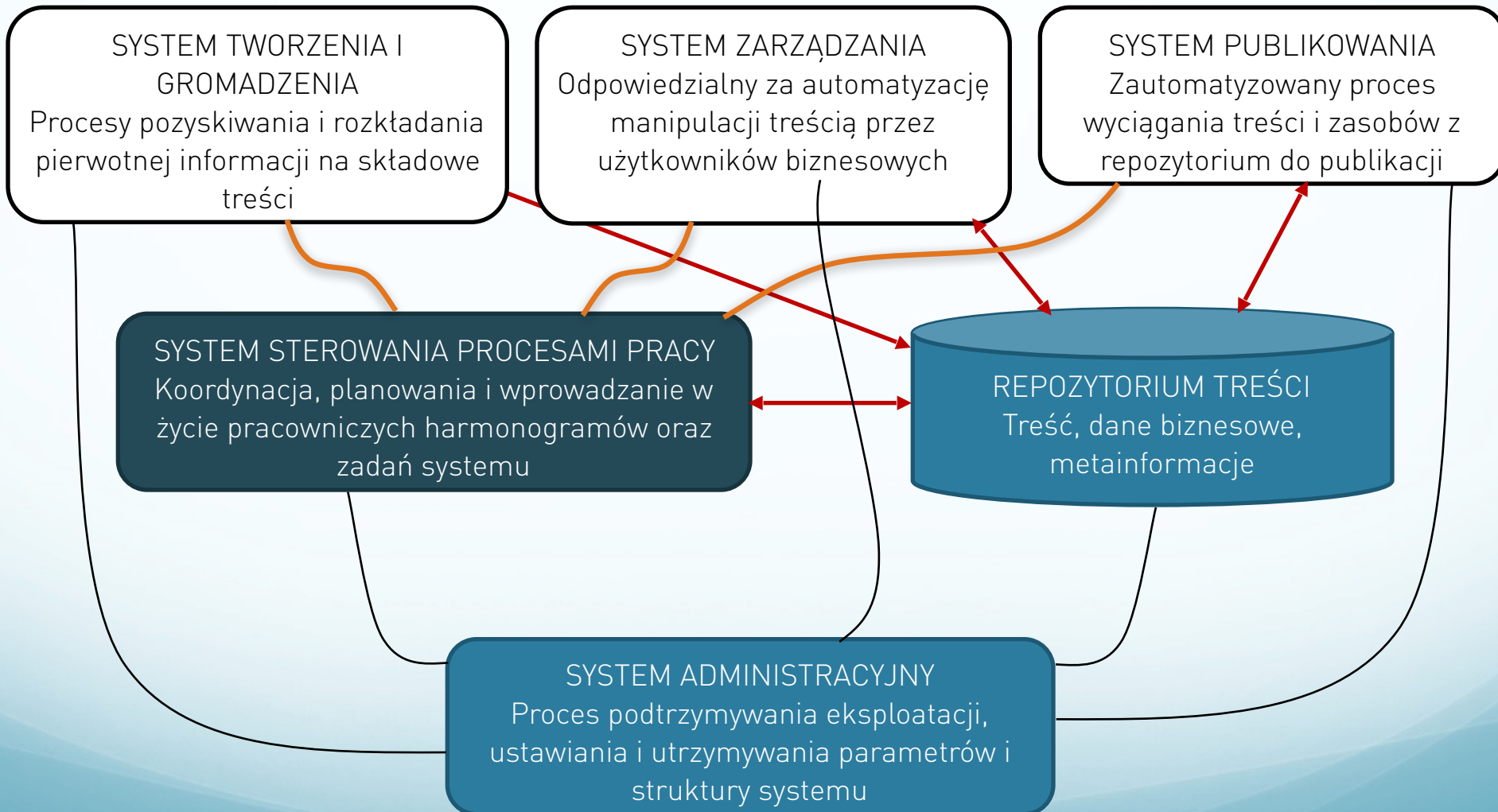


# Tematy związane z zarządzaniem treścią





# Główne komponenty CMS



# System tworzenia i gromadzenia treści

- Składa się z narzędzi, procedur oraz personelu, który jest zatrudniony w celu tworzenia i zbierania treści oraz wykonywania czynności redakcyjnych. Zadania:
  - Wytwarzanie treści "od zera": autorzy projektują, tworzą i poprawiają treść w wybranych przez siebie narzędziach.
  - Pozyskiwanie: dostosowywanie i redakcja treści z zewnętrznych źródeł.
  - Agregacja: formatowanie stylistyki informacji i ustalanie jej przeznaczenia: użyteczne składowe, meta-dane.
  - Konwersja: zmiany formatu i struktury informacji tak, aby spełniała ona wymagane standardy przechowywania treści; oddzielanie niepotrzebnych informacji np., nagłówek i stopki; odwzorowania tego formatu na wymagany standard, np. XML, który może być wprowadzony do systemu.
  - Usługi: są częścią logiki aplikacji oraz usług biznesowych dostarczonych przez CMS, wspomagających gromadzenie informacji oraz jej transformację. Usługi wspierają tworzenie, aktualizację i usuwanie składowych treści.

# System zarządzania

- Odpowiada za gromadzenie, przechowywanie, udostępnianie, pielęgnację i administrację składowych treści i innych zasobów informacji.
- Jest oparty na bazie danych treści, meta-informacji oraz danych biznesowych.
- Obejmuje procesy i narzędzia umożliwiające dostęp, aktualizację i administrowanie zgromadzoną informacją.
- Jest odpowiedzialny za bezpieczeństwo i autoryzację dostępu do treści.
- Jest odpowiedzialny za połączenia z innymi systemami.

# System sterowania procesami pracy

- Realizuje koordynację, planowanie i wprowadzanie w życie harmonogramów oraz zadań pracowników.
- Obejmuje narzędzia, procedury i ludzi zatrudnionych w celu zapewnienia skutecznych procesów zbierania, przechowywania i publikacji treści.
- System sterowania procesami pracy ma wpływ na system gromadzenia treści, system zarządzania oraz system publikowania.
- Każdy krok procesu, od wytwarzania po ostateczną publikację, powinien być możliwy do zamodelowania i śledzenia w obrębie tego samego systemu.
- Aspekty procesów pracy włączają: pracowników, zadania, czynności, standardowe procesy, narzędzia, czas, przepływ danych i dokumentów.

# System publikowania

- Jest odpowiedzialny za wyciąganie składowych treści i innych zasobów z repozytorium, formatowanie ich i automatyczne tworzenie z nich publikacji.
- Składa się z narzędzi, procedur i ludzi pobierających treść z repozytorium i tworzących publikacje.
- Powinien zawierać:
  - Szablony publikacji,
  - Kompletny język programowania,
  - Zależności pomiędzy treścią,
  - Dobrze zorganizowany system plików i katalogów,
  - Mechanizm ostatecznej publikacji.

# Procesy zarządzania treścią

- Włączają projektowanie, tworzenie, pozyskiwanie, recenzowanie, zatwierdzanie, konwersję, składowanie, testowanie i wdrożenie treści we wszystkich wymaganych miejscach.
  - Włączają pielęgnowanie, monitorowanie, uaktualnianie, wycofywanie i archiwizowanie treści.
  - Włączają komponenty raportujące i analityczne, celem świadomego usprawniania i poszerzania procesów zarządzania treścią.
- Wymagają jasnego zdefiniowania ról personelu oraz udokumentowanych procesów pracy dla wszystkich form treści.
- Mogą lecz nie muszą być wspomagane komputerowo.
  - Dla małych zastosowań wspomaganie jest często niepotrzebne.
  - Dla dużych zastosowań wspomaganie jest zazwyczaj niezbędne.

# Scenariusze i formy aplikacji zarządzania treścią

- Udostępnianie wiadomości (portale), np. internetowe gazety,
  - w tym wortale (vortals), czyli wiadomości ukierunkowane branżowo.
- Wyszukiwarki stron WWW (Yahoo, Altavista, Google, ...)
- Techniczne wspomaganie produktów danej firmy.
- B2C (Business-To-Customer): e-handel - sklepy internetowe.
- Portale wymiany informacji w danej dziedzinie, portale edukacyjne.
- B2B (Business-To-Business): e-biznes (portale biznesowe): transakcje, sprzedaż lub wymiana towarów i usług, w skali hurtowej.
- B2E (Business-To-Employee): wewnętrzne systemy internetowe lub Intranetowe do obsługi procesów biznesowych wewnątrz firmy.
- C2C (Customer-To-Customer): ogłoszenia drobne, aukcje, ...
- Portale korporacyjne (corporate portals) - organizują rozproszone i heterogeniczne zasoby i usługi informacyjne danej organizacji.
- Praca grupowa rozproszonych zespołów, wirtualne biura projektowe.
- .... wiele innych możliwości ....

# Funkcje wspólne dla wielu form i scenariuszy zarządzania treścią (1)

- **Projektowanie.** Zasadniczo nie odbiega od metod projektowania baz danych np. poprzez diagramy encja-związek lub UML.
- **Tworzenie.** Rola wykonywana przez autorów tekstu, fotografów, artystów grafików, producentów video, producentów dźwięku, specjalistów od reklamy i marketingu, prawników, lub kogokolwiek innego, kto produkuje oryginalny materiał przeznaczony dla użytkownika WWW.
- **Pozyskiwanie lub adoptowanie treści z istniejących źródeł.**
- **Klasyfikacja, indeksowanie.** Treść musi mieć przypisane cechy formalne (np. datę utworzenia, autora, itd.) oraz cechy klasyfikacji przedmiotowej (np. kategorię przedmiotową lub słowa kluczowe). Funkcja jest często określana jako wiązanie treści i metadanych.



# Funkcje wspólne dla wielu form i scenariuszy zarządzania treścią (2)

- **Recenzje i przeglądy.** Są wymagane dla wszystkich rodzajów udostępnianej treści.
- **Zatwierdzenie.** Formalne zatwierdzenie publikowanej treści jest niezbędnym składnikiem prawnej odpowiedzialności za treść.
- **Konwersja.** Tekst, grafika, dźwięk, i inne formy treści muszą być przystosowane do formatu najwygodniejszego lub obowiązującego w danym CMS, np. do formatu HTML lub XML.
- **Przechowywanie.** Treść jest zwykle przechowywana w plikach lub w bazie danych.
  - Dla większych zastosowań treść musi podlegać zarządzaniu konfiguracją (Software Configuration Management, SCM), w szczególności musi podlegać zarządzaniu wersjami oraz śledzeniu i kontrolowaniu zmian.

# Funkcje wspólne dla wielu form i scenariuszy zarządzania treścią (3)

- **Testowanie.** Może dotyczyć różnych aspektów:
  - błędnych lub nieaktualnych linków,
  - stron wolno ładujących się,
  - błędów w skryptach lub apletach, np. pętli,
  - błędów w komunikacji od klienta do serwera
- **Dojrzewanie.** Rodzaj testowania, polegający na weryfikacji kompletności i spójności większego zespołu treści, np. informacji o różnych aspektach nowej usługi.
- **Wdrożenie.** Obejmuje wszelkie fizyczne aspekty udostępnienia treści dla jej użytkowników, w tym replikacje treści na różnych serwerach.
- **Pielęgnacja, aktualizacja, zmiany.** Obserwowanie udostępnianej treści i reakcja na wszelkie sygnały i potrzeby zmian.

# Funkcje wspólne dla wielu form i scenariuszy zarządzania treścią (4)

- **Wycofywanie i archiwizacja.** Wycofanie może nastąpić z wielu powodów, np. utraty aktualności, utraty praw do treści, uatrakcyjnienie portalu nowszą treścią, niską frekwencją odwiedzania, itd. Przyjmuje się, że dowolna wycofywana treść podlega archiwizacji a/a.
- **Raporty i analizy.** Obejmuje różne formy raportów i analiz mających na celu lepszą obsługę użytkowników, zwiększenia atrakcyjności portalu, zbadania efektywności biznesowej, itd.
- **Ponowne użycie.** Wyodrębnienie i generalizacja pewnych elementów treści, metadanych, procesów, funkcji, szablonów formularzy, itd. jako udokumentowanych aktywów ponownego użycia w ramach danego repozytorium; opisywanie i propagowanie aktywów ponownego użycia wśród personelu.

# Klasyfikacja i przegląd CMS (1)

- Duże pakiety obejmujące funkcjonalnością wszystkie etapy i aspekty tworzenia systemów Internetowych.
  - Przykłady: V/6 Content Management Suite (Vignette), One-To-One Publishing (Broadvision), Content Server (Divine).
- Produkty o cechach podobnych jw., o mniejszych możliwościach integracji z istniejącymi systemami produkcyjnymi
  - Przykłady: Content Management Server (Microsoft), PVCS Content Manager (Merant), RedDot Solutions (RedDot), Mediasurface 3.5 (Mediasurface)
- Narzędzia, w których główny nacisk położono na zarządzanie dużymi repozytoriami dokumentów i wspomaganie pracy grupowej
  - Przykłady: Xpedio Content Management Suite (Stellent), 4I WCM Edition (Documentum), Panagon (FileNET)

# Klasyfikacja i przegląd CMS (2)

- Systemy, które służą do zarządzania cyklem wytwarzania elementów stanowiących treść serwisu (zagadnienia związane z rolami użytkowników, procesem prac)
  - Przykłady: TeamSite (Interwoven), CommonSpot Content Server (PaperThin)
- Narzędzia wspierające końcową fazę powstawania serwisu internetowego czyli jego publikację, personalizację itp.
  - Przykłady: WebLogic E-Business Platform (BEA), Dynamo e-business Platform (ATG), Oracle9iAS(Oracle)
- Systemy tworzone w ramach projektów „open-source”:
  - Przykłady: Content Management Framework (Zope), Arsdigita Community System (ArsDigita)

# Rodzaje rozwiązań CMS

Typ rozwiązania	Przykłady
System zarządzania dokumentami (DMS)	Documentum, Panagon 2000, ChangingPages
Systemy do publikowania w sieci gazet on-line, magazynów itp.	Eroom, Expressroom, Conversant, SlashDot, Frontier
E-business / E-Commerce	Vignette, Broadvision, ATG Dynamo, Open Market
Systemy do zarządzania wersjami	Content Management Studio, Interwoven TeamSite
Warstwa pośrednia (generowanie zawartości)	Enhydra, Midgard (PHP), Dmind DSM [ColdFusion], SiteGenesis, Obtrees.
„Wzorcowe” narzędzia zarządzania treścią	ACS, eGrail, Engenda, Mediasurface, WebSphere, NetObjects Fusion, Spectra, Xpedio, Zope
Procesory języka XML	Cocoon, Interwoven TeamSite Templating, Lychee, Rhythmix, Tamino, POET

# CMS jako katalizator rozwoju

- Dla wspomagania funkcji zarządzania treścią, dostawcy oprogramowania wytworzyli ogromną liczbę różnych narzędzi, zintegrowanych systemów, oraz rozszerzeń istniejących systemów.
  - Tradycyjna dziedzina zarządzania dokumentami została w dużym stopniu zdominowana przez funkcje CMS.
  - Pojawiły się obiektowe repozytoria do przechowywania treści dowolnego typu, w szczególności repozytoria XML.
  - Dostawcy systemów zarządzania bazami danych, tacy jak Oracle, CA, Sybase, Informix, IBM, przesunęli punkt ciężkości oferowanych SZBD z tradycyjnego zarządzania danymi na zarządzanie treścią, ze wspomaganiami tworzenia i podtrzymywania aplikacji Internetowych.
  - Znaczenia nabrały różnorodne formy procesów pracy (workflows) jako środka kontroli funkcji CMS.
  - Pojawiły się kompleksowe narzędzia typu "wszystko w jednym".

# Cechy CMS: procesy pracy i automatyzacja procesów biznesowych ost 3.10

- Zarządzanie treścią wymaga rutynowych usług znanych z procesów pracy (workflows), takich jak: śledzenie, przypisywanie ról i odpowiedzialności, zintegrowane bezpieczeństwo, zautomatyzowane zawiadamianie, monitorowanie populacji procesów.
- Systemy zarządzania procesami prac umożliwiają zdefiniowanie wielo-krokowych procesów włączających różnorodną treść, personel, oraz akcje takie jak wysłanie, recenzja, zatwierdzenie, itd.
- Systemy zarządzania procesami prac zapewniają automatyzację takich zadań jak:
  - ustalenie zespołów ludzkich i ról osób w zespołach
  - projektowanie procesów pracy
  - tworzenie i podtrzymywanie działania instancji procesów pracy.



# Cechy CMS: personalizacja

- Istotą personalizacji jest:
  - Rejestracja i autentyfikacja użytkowników aplikacji internetowej.
  - Dostosowanie się serwisu internetowego do indywidualnych preferencji użytkownika; np. jego preferencje tematyczne.
  - Przechowywanie i udostępnianie spersonifikowanych treści wprowadzanych przez użytkownika lub sparametryzowanych przez użytkownika; np. jego notatki, zakładki, kalendarz, terminarz zdarzeń, terminarz przypomnień, itd.
  - Przechowywanie historii odwiedzin serwisu przez użytkowników oraz transakcji lub zakupów, które oni dokonali.
  - Syntetyczne analizy i raporty dotyczące użytkowników mające na celu zwiększenie jakości i efektywności treści i usług oferowanych przez aplikację Internetową.

# Podejścia do personalizacji

- Środki techniczne:
  - Informacja dokładna o zarejestrowanych użytkownikach zgromadzona w bazie danych po stronie serwera
  - Informacja o konkretnym koncie i konkretnym komputerze na którym pracuje użytkownik, na podstawie tzw. ciasteczek (cookies).
    - Ciasteczka są plikami pamiętanymi przez komputer klienta, w których serwer może zapisać dowolną (niezbyt długą) informację.
    - Konkretny użytkownik jest nieznan, znana jest tylko jego "tożsamość" z dokładnością do konta i komputera.
    - Ciasteczka nie są dzielone pomiędzy różne portale, każdy portal ma swoje.
- Środki dostosowywania treści do profilu użytkownika:
  - Określanie profilu zainteresowań explicite przez użytkownika.
    - Wadą jest to, że on zwykle tego nie robi, a jeżeli robi, to niezbyt dokładnie.
  - Kolaboracyjne filtrowanie (collaborative filtering): tworzenie kategorii użytkowników i przypisywanie użytkownika do określonej kategorii na podstawie historii jego zachowania się na portalu (klikologii)..

# Cechy CMS: wyszukiwanie

- Sprawny mechanizm wyszukiwania informacji przez końcowych użytkowników jest czynnikiem powodzenia aplikacji internetowej.
- Wyszukiwanie oznacza konieczność klasyfikacji treści i inteligentnego jej zaindeksowania.
- Wyszukiwanie często odbywać się po cechach formalnych (data publikacji, autor, kategoria tematyczna, słowa kluczowe), które są niekiedy określane (w RDF) jako "metadane".
- Częściej wyszukiwanie odbywa się:
  - w pełnym tekście przechowywanych składników treści
  - poprzez asocjacje elementów treści z innymi elementami treści
- Klasyczne formy wyszukiwania (znane) z bibliotek są mało użyteczne.
- Konieczne są nowe paradygmaty, z reguły oparte o metafory graficzne.

# Cechy CMS: ontologia

- W filozofii: nauka o bytach, teoria bytu, opis charakteru i struktury rzeczywistości, specyfikacja konceptualizacji.
- W sztucznej inteligencji: formalna specyfikacja (przy użyciu logiki matematycznej) obiektów, pojęć i innych bytów, które istnieją w pewnej dziedzinie, oraz formalna specyfikacja związków, które pomiędzy tymi bytami zachodzą.
  - Podejście sztucznej inteligencji jest naiwne. Np. Giełda Papierów Wartościowych: wiele tysięcy stron aktów prawnych, zarządzeń, regulacji, itd. Kto to zapisze przy użyciu formuł rachunku predykatów?
- W biznesie (ontologia biznesowa, business ontology): wszystko to, co projektanci systemów informatycznych powinni wiedzieć o biznesie, aby poprawnie napisać aplikacje wspomagające ten biznes.
- Wiedza ta powinna być formalnie zapisana. "Formalnie" oznacza zwykle pewien standardowy i uzgodniony język, np. XML/RDF.

# Cechy CMS: metadane

- Ogólna definicja: są to dane o danych - co dane zawierają, jaką mają budowę, jakie jest ich znaczenie, jakim podlegają ograniczeniom, jak są zorganizowane, przechowywane, zabezpieczane, udostępniane, itd.
- Metadane są pewnym rozszerzeniem pojęcia schematu bazy danych, albo też pewną implementacją tego schematu w postaci katalogów.
- Metadane przykrywają także informację niezależną od treści samych danych, np. kiedy pewna dana została utworzona, w jakim jest formacie, kto jest jej autorem, do kiedy jest ważna, itd.
- Opisy danych zawarte w metadanych mają dwie podstawowe zalety:
  - Zawierają wspólne abstrakcje dotyczące reprezentacji danych, takie jak format; ogólnie "wyciągają przed nawias" wszystkie wspólne informacje, co redukuje znacznie objętość samych danych;
  - Reprezentują wiedzę dziedzinową (ontologię); umożliwiają wnioskowanie o danych, mogą być przez to użyte do redukcji dostępu do samych danych.

# Ontologia i metadane

- Głównym celem prac na biznesową ontologią jest standardyzacja następujących elementów:
  - Gramatyki opisów poszczególnych bytów,
  - Nazw i znaczeń nazw obowiązujących w ramach danego biznesu (np. co oznaczają słowa "autor", "klient", "instrument", "akcja", itd.),
  - Ograniczeń związanych z opisywanymi bytami,
  - Metadanych związanych z bytami (autor opisu, data stworzenia opisu, data ostatniej aktualizacji, itd.),
  - Dopuszczalnych operacji na bytach.
- W tym zakresie zapis ontologii jest pewną meta-bazą danych, w które ustala się zarówno strukturę samej bazy danych, jak i pewne dodatkowe informacje (meta-atrybuty) będące podstawą przetwarzania bazy danych.