

IBM Content Manager for Multiplatforms



Zarządzanie programem Information Integrator for Content

Wersja 8 wydanie 2

IBM Content Manager for Multiplatforms



Zarządzanie programem Information Integrator for Content

Wersja 8 wydanie 2

Uwaga

Przed użyciem podanych poniżej informacji oraz produktu, którego one dotyczą, należy zapoznać się z informacjami zawartymi w rozdziale "Uwagi" na stronie 135.

Wydanie drugie (Marzec 2003)

Niniejsze wydanie dotyczy programu IBM Information Portal for Multiplatforms wersja 8 wydanie 2 (numer produktu 5724-B43) i wszystkich nowszych wydań i modyfikacji, o ile w nowych wydaniach nie będzie stwierdzone inaczej.

Prawa autorskie do części produktu: Copyright © 1990-2000 ActionPoint, Inc. lub jego licencjobiorcy, 1299 Parkmoor Drive, San Jose, CA 95126 U.S.A. Wszelkie prawa zastrzeżone.

Outside In[®] Viewer Technology, ©1992-2000 Inso Corporation. Wszelkie prawa zastrzeżone.

© Copyright International Business Machines Corporation 1999, 2003. Wszelkie prawa zastrzeżone.

Spis treści

Informacje o tym podręczniku v

Do kogo adresowany jest ten podręcznik	v
Umiejętności wymagane od administratorów	vi
Wymagania, które muszą spełnić analitycy biznesowi lub projektanci procesów	vi
Gdzie można znaleźć więcej informacji	vi
Informacje zawarte w pakiecie	vi
Wsparcie techniczne dostępne w sieci WWW	vii
Jak wysłać swoją opinię	viii
Nowe opcje w programie EIP wersja 8.2	viii

Rozdział 1. Wprowadzenie do programu Enterprise Information Portal. 1

Wyszukiwanie informacji o klientach	1
Potrzeby	2
Rozwiązanie	2
Przegląd	2
Wprowadzenie do składników programu EIP	3

Rozdział 2. Klient administracyjny - wprowadzenie 9

Pierwsze kroki w programie klient administracyjny systemu 9	
Administrowanie programem EIP	9
Zarządzanie użytkownikami i grupami	9
Używanie narzędzi klienta administracyjnego	9
Wprowadzenie do uprawnień	11
Uprawnienia	11
Przełączanie widoku produktu i baz danych	12
Udoskonalenia w kliencie administracyjnym	12
Łączenie klienta administracyjnego z lokalną administracyjną bazą danych	13
Łączenie klienta administracyjnego ze zdalną administracyjną bazą danych	13
Krok 1 - skataloguj zdalną bazę danych przy użyciu programu DB2 Configuration Assistant	13
Krok 2 - użyj programu Server Configuration Utility	14
Krok 3 - testowanie połączenia ze zdalną bazą danych	16
Definiowanie typów dokumentów	16
Modyfikowanie pliku typu MIME (cmbcc2mime.ini)	16

Rozdział 3. Używanie narzędzi klienta administracyjnego. 19

Tworzenie wyszukiwania stowarzyszonego	19
Definiowanie serwerów	19
Wytyczne do definiowania serwerów	20
Praca z łącznikiem OnDemand: strojenie TCP/IP i gniazdek	25
Praca z łącznikiem wyszukiwania rozszerzonego.	25
Tworzenie jednostek stowarzyszonych.	25
Informacje o jednostkach stowarzyszonych	25
Korzystanie z kreatora tworzenia jednostek stowarzyszonych	27
Tworzenie stowarzyszonych indeksów tekstowych	27
Tworzenie szablonów wyszukiwania	27

Definiowanie szablonu wyszukiwania	28
Definiowanie kryteriów wyszukiwania.	28
Określanie ustawień wyszukiwania.	28
Przypisanie uprawnień	29

Rozdział 4. Zarządzanie dostępem użytkowników. 31

Tworzenie identyfikatorów i haseł użytkowników	31
Uprawnienia administracyjne bazy danych DB2	32
Łączenie się z bazą danych DB2 przy użyciu plików INI	32
Zmiana serwera bibliotecznego i hasła administratora systemu w menedżerze zasobów	33
Zmiana haseł dostępu do bazy danych	33
Importowanie użytkowników z katalogu LDAP	34
Wprowadzenie do uprawnień	34
Tworzenie zbiorów uprawnień	35
Tworzenie grup uprawnień	35
Przypisywanie użytkownikowi zbioru uprawnień	36
Przypisywanie użytkownikowi zbioru uprawnień nadawania	36
Przypisywanie użytkowników do menedżera zasobów	36
Przypisywanie użytkowników do kolekcji.	36
Tworzenie grup użytkowników	36
Tworzenie list kontroli dostępu	37
Przypisywanie zbiorów uprawnień do listy kontroli dostępu	37
Tworzenie domen	37
Administrowanie domenami.	38
Dostęp do domen	38
Przypisywanie użytkowników do domeny.	39
Przypisywanie grupy użytkowników do domeny.	39
Przypisywanie zbioru uprawnień do domeny	39
Przypisywanie menedżera zasobów do domeny	39
Przypisywanie kolekcji do domeny.	39
Przenoszenie użytkownika do innej domeny	40
Przenoszenie grupy użytkowników do innej domeny	40
Przenoszenie menedżera zasobów do innej domeny	40
Przenoszenie kolekcji do innej domeny	41
Przenoszenie zbioru uprawnień do innej domeny.	41
Przenoszenie listy kontroli dostępu do innej domeny	41

Rozdział 5. Zarządzanie eksploracją informacji 43

Co to jest Eksploracja informacji?	43
Usługi eksploracji informacji w produkcie Enterprise Information Portal	43
Składniki usług eksploracji informacji	44
Używanie eksploracji informacji w środowisku biznesowym	46
Przykład zastosowania produktu Eksploracja informacji	48
Obsługiwane języki i formaty	51
Koncepcja	51
Architektura systemu	52
Koncepcje eksploracji informacji	53

Narzędzia programu Eksploracja informacji	54
Interfejs programisty	61
Pierwsze kroki	62
Tworzenie systematyki	63
Instalowanie programu Information Structuring Tool	63
Pierwsze kroki	63
Prawa dostępu	64
Definiowanie systematyki	64
Wybieranie dokumentów szkoleniowych	66
Wysyłanie dokumentów szkoleniowych	66
Wartościowanie modelu klasyfikacji	69
Szkolenie katalogu	72
Strojenie wydajności	73
Używanie programu IBM Web Crawler	73
Możliwości programu IBM Web Crawler	74
Konfigurowanie i uruchamianie programu IBM Web Crawler dla sieci WWW	74
Plik konfiguracyjny programu IBM Web Crawler	77
Protokołowanie w programie IBM Web Crawler	85
Rozwiązywanie problemów	86
Wybór programów streszczających	87
IBM Web Crawler dla Notes	89
Wykluczanie programu IBM Web Crawler z serwera	92

Rozdział 6. Wprowadzenie do sterowania przepływem pracy 95

Pojęcie przepływu pracy	95
Jak korzystać z funkcji przepływu pracy	95
Synchronizowanie identyfikatorów użytkowników i grup	96
Ponowne instalowanie serwera EIP z włączoną opcją przepływu pracy	97
Aktualizowanie identyfikatorów użytkowników i grup między MQSeries Workflow i bazą danych EIP	97
Planowanie przepływu pracy	98
Informacje do przetworzenia	99
W jaki sposób informacje są przetwarzane	99
Czynności, które należy wykonać	99
Przepływ informacji w procesie	99
Jak wszystkie elementy działają razem	100
Korzystanie ze składników przepływu pracy programu Enterprise Information Portal	100
Używanie programu budującego przepływ pracy	100
Używanie usług przepływu pracy	101
Definiowanie list prac	101
Definiowanie list czynności	102
Tworzenie przepływu pracy	102
Włączanie programu budującego przepływ pracy	102
Uruchamianie serwera MQSeries Workflow	102

Rozdział 7. Przykładowe pliki programu IBM Web Crawler 105

Przykładowy plik config-sample2.xml	105
Analiza przykładowego pliku protokołu programu IBM Web Crawler	107

Rozdział 8. Używane wyszukiwania tekstowego i QBIC 111

Wyszukiwanie dokumentów za pomocą mechanizmu wyszukiwania tekstowego	111
Włączanie serwera wyszukiwania tekstowego	111
Wyszukiwanie obrazów za pomocą języka QBIC	111
Wprowadzenie do wyszukiwania obrazów	112
Konfigurowanie funkcji wyszukiwania obrazów	112
Ładowanie i indeksowanie przykładowych danych	115
Przed załadowaniem przykładowych danych	115
Tworzenie indeksu wyszukiwania tekstowego	115
Tworzenie bazy danych, katalogu i funkcji wyszukiwania obrazów	116
Uruchamianie programu ładującego	117
Indeksowanie przykładowych danych tekstowych	118

Rozdział 9. Formaty dokumentów 121

Formaty dokumentów programu eksploracji informacji	121
Przetwarzanie tekstu: Ogólne	121
Przetwarzanie tekstu: DOS	121
Przetwarzanie tekstu: Międzynarodowe	122
Przetwarzanie tekstu: Windows	122
Przetwarzanie tekstu: Macintosh	123
Arkusze kalkulacyjne	123
Bazy danych	123
Pliki graficzne	124
Profesjonalne formaty graficzne	126
Formaty prezentacyjne	126
Formaty skompresowane i zakodowane	126
Inne	127

Rozdział 10. Zarządzanie prawami 129

Ochrona własności intelektualnej	129
Korzystanie z technologii znakowania	130
Oznakowanie widoczne	131
Oznakowanie niewidoczne	131

Rozdział 11. Funkcje dostępności 133

Wprowadzanie danych z klawiatury i nawigowanie	133
Funkcje wyświetlania	133
Kompatybilność z technologiami asysty	133
Dostępna dokumentacja	134

Uwagi. 135

Znaki towarowe	137
--------------------------	-----

Glosariusz. 139

Indeks 147

Informacje o tym podręczniku

Niniejsza publikacja jest wprowadzeniem do wszystkich podstawowych pojęć, których poznanie jest niezbędne do administrowania systemem Enterprise Information Portal (EIP). Ponieważ w programie EIP dostępnych jest kilka składników, które mogą być zarządzane za pomocą klienta administracyjnego i ponieważ używając programu EIP można uzyskać dostęp do innych produktów, niniejsza publikacja nie jest typowym podręcznikiem dla administratora. W tym dokumencie opisano:

- Jak używać programu EIP w celu zaspokojenia potrzeb firmy.
- Jak uzyskiwać dostęp i używać klienta administracyjnego.
- Jak zarządzać dostępem użytkowników.
- Jak używać programu EIP do wyszukiwania danych na wielu serwerach danych, w tym danych strukturalnych przechowywanych w relacyjnych bazach danych, danych niestukturalnych lub multimedialnych, a także dokumentów tekstowych.
- Jak projektować, wdrażać i zarządzać procesem przepływu pracy.

Do kogo adresowany jest ten podręcznik

Podręcznik ten ma za zadanie pomagać administratorom programu EIP w wykonywaniu następujących zadań:

Administrowanie systemem

Obejmuje administrowanie bazą danych, serwerem oraz siecią.

Zarządzanie użytkownikami

Obejmuje definiowanie i udzielanie dostępu użytkownikom i ich grupom oraz zarządzanie listami kontroli dostępu.

Wyszukiwanie stowarzyszone

Definiowanie i używanie szablonów wyszukiwania stowarzyszonego w celu uzyskania danych z systemu zarządzania danymi.

Eksploracja informacji

Wyodrębnianie informacji z dokumentów, dzielenie informacji na kategorie i przeszukiwanie.

Indeksowanie sieci WWW

Używanie programu IBM Web Crawler do przeszukiwania i importowania danych z sieci WWW.

Wyszukiwanie tekstowe

Używanie programu IBM DB2 TIE lub IBM Text Search Engine (tylko dla produktu Content Manager w wersji 7.1 i starszych) do wyszukiwania i indeksowania dokumentów.

Wyszukiwanie obrazów

Używanie programu Content Manager w wersji 7.1 (i starszych) do wyszukiwania obrazów.

Zarządzanie przepływem pracy

Używanie narzędzi przepływu pracy dostarczanych z programem EIP w celu zarządzania przepływem pracy w przedsiębiorstwie.

Umiejętności wymagane od administratorów

W zależności od wykonywanych zadań czytelnik musi znać:

- protokoły ochrony dostępu użytkowników
- Systemy operacyjne Windows NT, Windows XP, Windows 2000, AIX, i Solaris.
- zagadnienia administrowania siecią
- modele danych serwerów danych w projektowanym systemie zarządzania danymi
- administrowanie bazą danych
- stosowanie praktycznej wiedzy o danych i kryteriach wyszukiwania przy tworzeniu szablonów wyszukiwania
- techniki eksploracji informacji oraz służące do tego celu narzędzia
- zasady przepływu pracy
- procesy biznesowe, które mają być obsługiwane przez program EIP

Wymagania, które muszą spełnić analitycy biznesowi lub projektanci procesów

Analitycy biznesowi i projektanci procesów również znajdą w niniejszym podręczniku informacje, które mogą być przydatne podczas definiowania modelu przepływu pracy EIP w przedsiębiorstwie.

Aby używać programu budującego Enterprise Information Portal, trzeba:

- znać wymagania personelu oraz programy i struktury danych używane w procesach biznesowych przedsiębiorstwa,
- podejmować decyzje dotyczące procesów biznesowych lub procesów przepływu pracy w przedsiębiorstwie.

Gdzie można znaleźć więcej informacji

W dostarczonym pakiecie zawarty jest kompletny zestaw informacji pomocnych w planowaniu, instalowaniu, administrowaniu i używaniu systemu. Dokumentacja produktu oraz obsługa dostępne są przez sieć WWW.

Informacje zawarte w pakiecie

Wraz z produktem dostarczane jest Centrum informacyjne oraz dokumentacja w przenośnym formacie PDF.

Centrum informacyjne

Podczas instalacji produktu można zainstalować dostarczane wraz z nim Centrum informacyjne. Informacje dotyczące instalowania Centrum informacyjnego zawiera sekcja *Planning and Installing Your Content Management System*.

Centrum informacyjne zawiera dokumentację produktów Content Manager, Enterprise Information Portal i IBM Content Manager VideoCharger for Multiplatforms. Informacje są podzielone na tematy i posegregowane według produktów i zadań (np. administrowanie). Oprócz dostarczanego mechanizmu nawigacji i indeksów, opcja wyszukiwania dodaje możliwość pobierania danych.

Publikacje w formacie PDF

Pliki PDF można przeglądać w trybie online za pomocą przeglądarki Adobe Acrobat Reader w używanym systemie operacyjnym. Program Acrobat Reader można pobrać z serwisu WWW firmy Adobe pod adresem <http://www.adobe.com>.

Tabela 1 zawiera spis publikacji dotyczących produktu Content Manager zawartych w IBM Content Manager for Multiplatforms.

Tabela 1. Publikacje dotyczące programu Content Manager

Nazwa pliku	Tytuł	Numer publikacji
install	<i>Planning and Installing Your Content Management System¹</i>	GC27-1332-01
migrate	<i>Migracja do programu Content Manager wersja 8</i>	SC85-0068-01
sysadmin	<i>System Administration Guide</i>	SC27-1335-01

Do każdego zamówionego programu IBM Content Manager for Multiplatforms dodawany jest pakiet IBM Enterprise Information Portal for Multiplatforms. Można go także zamówić oddzielnie. Tabela 2 przedstawia publikacje dotyczące produktu Enterprise Information Portal, które są w nim zawarte.

Tabela 2. Publikacje dotyczące programu Enterprise Information Portal

Nazwa pliku	Tytuł	Numer publikacji
apgwork	<i>Application Programming Guide for Windows¹</i>	SC27-1347-01
ecliinst	<i>Instalowanie, konfigurowanie i zarządzanie programem eClient</i>	SC85-0066-02
eipinst	<i>Planning and Installing IBM Information Integrator for Content</i>	GC85-0065-01
eipmanag	<i>Zarządzanie programem IBM Information Integrator for Content</i>	SC85-0067-01
messcode	<i>Komunikaty i kody²</i>	SC85-0069-01

Uwagi:

1. Publikacja *Application Programming Guide for Windows* zawiera informacje na temat tworzenia programowania dla Content Manager i Enterprise Information Portal.
2. Publikacja *Komunikaty i kody* zawiera komunikaty i kody produktów Content Manager i Enterprise Information Portal.

Wsparcie techniczne dostępne w sieci WWW

Wsparcie techniczne produktu dostępne jest w sieci WWW. Kliknij odsyłacz **Support (Obsługa)** na następujących stronach WWW:

www.ibm.com/software/data/cm/

www.ibm.com/software/data/eip/

Dokumentacja dostarczana jest w wersji elektronicznej wraz z produktem. Aby uzyskać dostęp do dokumentacji produktu w sieci WWW, kliknij **Library (Biblioteka)** na stronie WWW produktu.

Interfejs dokumentacji w języku HTML o nazwie Enterprise Documentation Online (EDO) również jest dostępny w sieci WWW. Obecnie zawiera odniesienia do interfejsu API. Dostęp do EDO możliwy jest za pomocą strony WWW Enterprise Information Portal Library.

Jak wysłać swoją opinię

Opinia użytkowników pomaga firmie IBM dostarczać informacji o wysokiej jakości. Wszelkie komentarze na temat publikacji lub produktów Content Manager i Enterprise Information Portal są mile widziane. Można przysyłać je na kilka sposobów:

- Za pomocą sieci WWW. Otwórz stronę IBM Data Management Online Reader's Comment Form (RCF):

www.ibm.com/software/data/rcf

Za pomocą tej strony można wysyłać komentarze.

- Za pomocą poczty elektronicznej wysłanej pod adres comments@vnet.ibm.com. Podaj nazwę produktu, numer wersji oraz numer publikacji. Jeśli komentowany jest określony tekst, podaj jego położenie, np. tytuł rozdziału, numer tabeli lub strony.

Nowe opcje w programie EIP wersja 8.2

W produkcie wprowadzono następujące zmiany:

Obsługa systemów Sun Solaris

W systemach Solaris można instalować łączniki, opcje i bazy danych.

Wspólne administrowanie systemem

Pojedyncza aplikacja kliencka zapewnia oddzielny dostęp do funkcji administracyjnych produktów Content Manager i Enterprise Information Portal.

Nowe łączniki

- Łącznik ICM dla produktu Content Manager wersja 8 wydanie 1 zapewnia możliwość wykorzystania zaawansowanych opcji przechowywania dokumentów.
- Nowy łącznik C++ Extended Search wersja 3.7 działa w systemie AIX.

Ulepszone łączniki

- Parametryczne wyszukiwanie tekstowe jest obsługiwane z warstwy stowarzyszonej i poprzez bezpośrednie połączenie Extended Search.
- Udoskonalenia funkcjonalne i większa wydajność łącznika OnDemand, w tym:
 - modyfikacje struktury DDO produktu OnDemand,
 - obsługa wyszukiwania asynchronicznego.

Nowe usługi eksploracji informacji

- Wyodrębnianie funkcji
- klastrowanie
- identyfikacja języka

IBM Web Crawler

Program IBM Web Crawler jest opcją umożliwiającą użytkownikom wyszukiwanie i streszczanie informacji, znajdujących się w sieci WWW i w bazach danych Lotus Notes.

Rozszerzenia przepływu pracy

Systemy AIX i Solaris obecnie w pełni obsługują przepływ pracy. Program budujący przepływ Pracy, funkcje API i komponent JavaBeans zwiększają funkcjonalność i użyteczność przepływu pracy.

Centrum informacyjne

Korzystające z przeglądarki WWW Centrum informacyjne zawiera dokumentację produktów Content Manager, Enterprise Information Portal i IBM Content Manager VideoCharger for Multiplatforms. Informacje są podzielone na tematy i

posegregowane według produktów i zadań (np. administrowanie). Oprócz dostarczanego mechanizmu nawigacji i indeksów, opcja wyszukiwania dodaje możliwość pobierania danych.

Dostępność

Funkcje dostępności są pomocne dla użytkowników niepełnosprawnych, na przykład z ograniczeniami związanymi z poruszaniem się lub wadami wzroku, i umożliwiają efektywne używanie oprogramowania. Najważniejsze funkcje dostępności w tym produkcie obejmują:

- Możliwość obsługi programu za pomocą klawiatury bez używania myszy.
- Obsługę zaawansowanych właściwości wyświetlania.
- Opcje powiązania ze sobą alarmów dźwiękowych i graficznych.
- Kompatybilność z technologiami asysty.
- Kompatybilność z funkcjami dostępności systemu operacyjnego.
- Przyjazne formaty dokumentacji.

Rozdział 1. Wprowadzenie do programu Enterprise Information Portal

Wiele przedsiębiorstw przetwarzających dokumenty w formie papierowej, takich jak firmy ubezpieczeniowe i instytucje finansowe, ma do czynienia z dużymi zbiorami informacji biznesowych. W wielu gałęziach przemysłu istnieje zapotrzebowanie na rozwiązanie umożliwiające dostęp do tego typu informacji i zarządzanie nimi.

Serwer danych to system programowy, w którym przechowywane są dane multimedialne, formularze, dokumenty, dane biznesowe oraz metadane, które umożliwiają pracownikom ich przetwarzanie. Jeśli nie ma możliwości efektywnego połączenia różnych serwerów danych, firma traci czas i pieniądze na powielanie informacji lub szkolenie pracowników w zakresie wyszukiwania informacji w różnych miejscach.

Enterprise Information Portal to przoduający produkt umożliwiający dostęp do wszystkich zasobów przedsiębiorstwa z pojedynczej stacji roboczej. Program EIP pomaga wydobywać jak najwięcej z posiadanych informacji i zbiorów multimedialnych przez połączenie różnych serwerów danych za pośrednictwem pojedynczego klienta. Za pomocą klienta programu EIP użytkownicy mogą szybko i jednocześnie uzyskiwać dostęp do wszystkich połączonych serwerów danych. Użytkownicy mogą także przeprowadzać eksplorację informacji i wykonywać zaawansowane wyszukiwanie w serwerach danych, obejmujące sieci WWW i intranet. Mogą także wykonywać czynności związane z przepływem pracy w ramach zdefiniowanych procesów biznesowych.

Program EIP umożliwia dostosowywanie aplikacji do indywidualnych potrzeb przedsiębiorstwa. Za pomocą pakietów narzędziowych programu EIP programiści mogą tworzyć aplikacje graficzne i sieciowe.

W tej sekcji umieszczono przegląd programu EIP. Funkcje programu EIP zostały zademonstrowane na przykładzie fikcyjnej firmy ubezpieczeniowej XYZ.

Wyszukiwanie informacji o klientach

Firma Ubezpieczenia XYZ, duża firma zajmująca się ubezpieczeniami mienia i ubezpieczeniami od następstw nieszczęśliwych wypadków, posiada obszerny zbiór zdjęć, roszczeń, polis, notatek, raportów ekspertów i innych dokumentów biznesowych.

W firmie w formie plików archiwalnych Lotus Domino.Doc są przechowywane wszystkie pisma wysyłane do posiadaczy polis, łącznie z elektronicznymi formularzami medycznymi i formularzami wycen. Wszystkie deklaracje polis, powiadomienia i faktury są archiwizowane w systemie Content Manager OnDemand w celu ich długoterminowego przechowywania z możliwością szybkiego dostępu. Wszystkie formularze roszczeń, fotografie i listy otrzymane od posiadaczy polis są w firmie XYZ przechowywane w folderze systemu Content Manager iSeries. Raporty ekspertów są przechowywane w katalogu DB2 Universal Database (DB2 UDB) Data Warehouse Center Information Catalog Manager. Firma XYZ posiada także różnego rodzaju dane multimedialne, takie jak wysokiej jakości obrazy wykorzystywane w celach reklamowych, prezentacyjnych oraz stosowane w nowych działach biznesowych. Te dane są przechowywane w systemie Content Manager, co pozwala na ich współużytkowanie. Inne informacje, na przykład dane o procedurach stosowanych w firmie, są w firmie XYZ również przechowywane w sieci intranet.

Potrzeby

Nie można przeprowadzać obsługi roszczeń, rozmów z klientami ani ogólnej obsługi posiadaczy polis, posługując się zawartością tylko jednego serwera, ponieważ pracownicy firmy XYZ potrzebują dostępu do wszystkich informacji o kliencie. Aby obsługa klientów była możliwa, pracownicy muszą mieć jednoczesny dostęp do różnych serwerów danych. Firmie Ubezpieczenia XYZ potrzebne jest rozwiązanie, które umożliwi połączenie serwerów danych z firmową siecią intranet w celu wyszukiwania i pobierania informacji. Istnieje również potrzeba rozszerzenia wykorzystywania komputerowego wspomaganie przepływu pracy.

Wielu różnych pracowników, od urzędników do agentów, potrzebuje dostępu do dokumentów. Niektóre pozycje przechowywane w firmie XYZ muszą być zabezpieczone przed nieautoryzowanym dostępem, natomiast inne powinny być dostępne bez ograniczeń. Wymagany jest także łatwy w użyciu interfejs użytkownika, aby wyeliminować potrzebę szkolenia pracowników.

Rozwiązanie

Firma Ubezpieczenia XYZ wdraża system EIP, ponieważ stosowane w nim techniki wyszukiwania informacji umożliwiają jej pracownikom łączenie i przeszukiwanie wszystkich firmowych serwerów w celu uzyskania potrzebnych danych. Gdy pracownik działu obsługi klienta firmy XYZ przyjmuje zgłoszenie, przy użyciu jednej operacji wyszukiwania stowarzyszonego uzyskuje wszystkie potrzebne informacje o posiadaczu polisy.

W firmie XYZ usługa pozyskiwania informacji programu EIP jest także używana do wyszukiwania informacji w celu ich pobrania z sieci intranet firmy. Istnieje również potrzeba rozszerzenia wykorzystywania procesów przepływu pracy.

Przegląd

Program EIP to wszechstronny produkt składający się z wielu współpracujących ze sobą składników, które zapewniają rozwiązanie idealnie dopasowane do indywidualnych potrzeb przedsiębiorstwa. Wykorzystujący architekturę wielowarstwową program EIP zawiera klienta administracyjnego przeznaczonego do zarządzania wyszukiwaniami, przykładowe programy klienckie do przeprowadzania wyszukiwań i złącza do nawiązywania połączeń z różnymi serwerami danych, takimi jak IBM Content Manager, Content Manager ImagePlus for OS/390, Content Manager OnDemand, Lotus Domino.Doc, DB2 Universal Database, DB2 DataJoiner i DB2 Data Warehouse Center Information Catalog Manager. Użytkownicy mogą tworzyć nowe łączniki dla innych serwerów danych za pomocą pakietu narzędzi do programowania łączników.

Architektura systemu EIP umożliwia aplikacjom klienckim przeprowadzanie pojedynczych wyszukiwań na jednym lub wielu serwerach danych. Aby przeprowadzić wyszukiwanie, aplikacja kliencka używa szablonów wyszukiwania zdefiniowanych przez administratora EIP.

Wykorzystując szablony wyszukiwania, klient wykonuje operację *wyszukiwania stowarzyszonego* równocześnie na wszystkich serwerach danych, których atrybuty rodzime zostały odwzorowane na atrybuty stowarzyszone używane przez szablon wyszukiwania. Szablony wyszukiwań systemu EIP zawierają kryteria wyszukiwania, które odwołują się do atrybutów stowarzyszonych odwzorowywanych na rodzime atrybuty każdego z serwerów danych. Szablony wyszukiwania są tworzone przez administratora. EIP zapewnia łączniki, umożliwiające wyszukiwanie i dostęp do danych przechowywanych na wielu serwerach danych. Serwery danych zwracają klientowi obiekty danych.

Architektura systemu EIP posiada następujące zalety:

- Pojedyncze zapytanie umożliwia dostęp do wielu różnych serwerów danych, obsługujących transakcje typu e-business i aplikacje związane z obsługą klienta.
- Możliwość eksploracji informacji z wielu serwerów danych, a także z sieci WWW.
- Dostęp procesu przepływu pracy do danych znajdujących się na wielu heterogenicznych serwerach danych.
- Możliwość tworzenia aplikacji klienckich, które są niezależne od miejsca przechowywania danych na dowolnym serwerze danych, wynikająca z rozdzielenia aplikacji klienckich, indeksów i danych.

Wprowadzenie do składników programu EIP

W tej sekcji został przedstawiony opis każdego ze składników systemu EIP oraz opis opcji instalacyjnych.

Tabela 3 na stronie 3 zawiera listę składników oraz zgodnych z nimi systemów operacyjnych.

Tabela 3. Zgodność systemów operacyjnych ze składnikami EIP

Składnik	Windows	AIX	Solaris	Notes
Administracyjna baza danych	tak	tak	tak	Database obejmuje funkcję programu budującego przepływ pracy
Klient administracyjny	tak	nie	nie	Klient może łączyć się z bazami danych zainstalowanymi w systemach Windows, AIX lub Solaris.
Łączniki	tak	tak	tak	
Eksploracja informacji	tak	tak	tak	
IBM Web Crawler	tak	tak	tak	
Klient wyszukiwania tekstowego	tak	tak	tak	
Klient wyszukiwania obrazów	tak	tak	tak	

Tabela 3. Zgodność systemów operacyjnych ze składnikami EIP (kontynuacja)

Składnik	Windows	AIX	Solaris	Notes
Pakiet narzędzi dla programistów łączników i przykłady	tak	tak	tak	<ul style="list-style-type: none"> Wersja dla systemu Windows obejmuje kod źródłowy umożliwiający skompilowanie przykładowego klienta. W systemie AIX nie jest instalowany przykładowy kod klienta. Przykłady przepływu pracy i funkcje API są instalowane razem z przykładowym łącznikiem stowarzyszonym.
Przeglądarka	tak	nie	nie	Instaluje klienta i przeglądarkę OnDemand.
Centrum informacyjne	tak	tak	tak	

Administrowanie

Składnik administracyjny zawiera administracyjną bazę danych i klienta administracyjnego. Zainstalowanie administracyjnej bazy danych powoduje także zainstalowanie opcji przepływu pracy.

Administracyjna baza danych: Administracyjna baza danych DB2 służy do zarządzania informacjami na temat użytkowników EIP i ich grup, poziomów uprawnień, identyfikatorów i haseł oraz innych informacji. Dostarcza ona także funkcji obsługi przepływu pracy i, opcjonalnie, eksploracji informacji. Istnieje możliwość zainstalowania wielu baz danych. Każda z nich obsługuje przepływ pracy programu EIP. Jeżeli używany jest system Content Manager wersja 8, można dodać tabele EIP do istniejącej bazy danych serwera bibliotecznego tego systemu. Baza danych może być współużytkowana, ponieważ serwer biblieczny zawiera wszystkie informacje wymagane przez produkt EIP.

Klient administracyjny: Klienta administracyjnego można zainstalować wyłącznie w systemach Windows. Istnieje możliwość zainstalowania wielu klientów. Jeśli zainstalowany jest system Content Manager w wersji 8, programami EIP i Content Manager wersja 8 można administrować z tego samego klienta.

Klient dostarcza interfejs, który pozwala administratorowi na:

- Definiowanie serwerów danych do wyszukiwania stowarzyszonego.
- Identyfikowanie jednostek i atrybutów rodzimych na serwerach danych i odwzorowywania ich na jednostki stowarzyszone.
- Zarządzanie spisem zasobów elementów rodzimych i atrybutów dla wszystkich zdefiniowanych przez użytkownika serwerów danych.
- Tworzenie szablonów wyszukiwania.

- Definiowanie użytkowników i grup użytkowników.
- Przypisywanie użytkownikom uprawnień i zbiorów uprawnień.
- Definiowanie dostępu do szablonów wyszukiwania i ustawianie warunków dla akcji, które mogą być wykonane przez użytkownika na wyszukanych danych.
- Projektowanie i zarządzanie procesem przepływu pracy.

Łączniki

Łączniki stanowią interfejs komunikacyjny pomiędzy klientami EIP, serwerami danych i administracyjną bazą danych. Łączniki serwera danych, takie jak łącznik Content Manager w wersji 7.1, dostarczają funkcji, dzięki którym program EIP może zalogować się do serwera, wyszukać informacji i zwrócić je do klienta administracyjnego lub klienta uruchomionego przez użytkownika końcowego. Łącznik stowarzyszony łączy klienta administracyjnego z administracyjną bazą danych.

Program EIP zawiera następujące łączniki:

- Łącznik stowarzyszony łączący klienta EIP z administracyjną bazą danych.
- Łącznik relacyjnej bazy danych dla DB2 7.1, sterownik JDBC 1.3 (tylko Java), ODBC 3.0 (tylko C++), DataJoiner 2.1.1.
- Łącznik Content Manager dla serwerów Content Manager w wersji 7.1.
- Łącznik Content Manager dla serwerów Content Manager w wersji 8.2.
- Łącznik Content Manager OnDemand dla Content Manager OnDemand w wersji 7.1.
- Content Manager dla VisualInfo dla 400 wersja 4.3 i wersja 5.1.
- Łącznik Content Manager ImagePlus for OS/390 dla ImagePlus/390 Folder Application Facility w wersji 3.1, Image Plus/390 ODM w wersji 3.1.
- Łącznik Lotus Domino.Doc dla Domino.Doc w wersji 3.0a, Desktop Enabler Version 3.0a.
- Łącznik wyszukiwania rozszerzonego dla wersji 3.7.
- Łącznik Information Catalog Manager dla DB2. Universal Database Visual Warehouse wersja 5.2, DB2 Universal Database wersja 7.2.

Funkcje

EIP ma cztery opcje dodatkowe.

Eksploracja informacji

Element ten udostępnia usługi lingwistyczne umożliwiające wyszukiwanie informacji ukrytych w dokumentach tekstowych znajdujących się na serwerach danych. Podczas przetwarzania dokumentu tekstowego tworzone są metadane, które mogą być uogólniane, kategoryzowane i przeszukiwane. Wymaganiem wstępnym dla tej funkcji jest produkt WebSphere Application Server 4.0 (wersja standardowa lub zaawansowana). W dodatku podobne dokumenty można łączyć, wyodrębniać dane z dokumentów, np. nazwiska osób lub nazwy firm oraz określać wersję językową dokumentu.

Klient wyszukiwania obrazów

Klient ten stanowi interfejs umożliwiający dostęp do funkcji wyszukiwania obrazów i administrowania nią na serwerze danych Content Manager wersja 7.

Klient wyszukiwania tekstowego

Klient ten stanowi interfejs umożliwiający dostęp do funkcji wyszukiwania tekstowego i administrowania nią na serwerze wyszukiwania tekstowego.

IBM Web Crawler

Web Crawler jest programem napisanym w języku Java, który służy do pobierania i eksploracji danych. Program ten może przeszukiwać sieci typu intranet, ekstranet, bazy danych Lotus Notes (bezpośrednio lub za pomocą serwera Domino), lokalne systemy plików i zbiory FTP.

Web Crawler może przeszukiwać dane i tekst zapisany w wielu formatach. Na przykład dane HTML mogą być przeszukiwane według adresu URL, tytułu, zawartości, daty modyfikacji i metadanych, takich jak autor, słowa kluczowe, opis. Użytkownicy mogą wybierać opcje wyszukiwania ze zbioru opcji dostępnych dla określonego typu danych. Zawartość lub przeszukiwane metadane są zapisywane na lokalnym dysku.

Przeglądarka danych

Zainstalowanie przeglądarki OnDemand również spowoduje zainstalowanie klienta OnDemand i innych plików wymaganych do przeglądania dokumentów pobieranych z serwera OnDemand.

Pakiety narzędzi i przykłady łączników

Produkt EIP zawiera pakiet narzędzi do tworzenia łączników, w którym znajdują się przykładowe programy, które można wykorzystać do eksperymentowania z różnymi funkcjami EIP, takimi jak:

- łączenie się i odłączanie od serwerów danych;
- wykonywanie przykładowych zapytań SQL i innych zapytań na serwerach danych;
- określanie typów MIME dla serwerów danych itd.

Pakiet narzędzi do programowania łączników dla Windows: Aby zainstalować pakiet narzędzi do programowania łączników i programów przykładowych na serwerach z systemem Windows, wybierz typ komputera stacja robocza programisty. Następnie wybierz składnik Pakiet narzędzi dla programistów łączników i przykłady. Możesz zainstalować przykładowe programy dla wszystkich łączników lub wybrać pojedyncze programy przykładowe, które odpowiadają zainstalowanym łącznikom.

Na serwerach Windows przykładowe programy dla pakietu narzędzi do tworzenia łączników są zorganizowane w następujący sposób:

```
c:\CMBROOT\SAMPLES\activex\xx
c:\CMBROOT\SAMPLES\cpp\xx
c:\CMBROOT\SAMPLES\java\xx
c:\CMBROOT\SAMPLES\jsp\xx
c:\CMBROOT\SAMPLES\server\xx
```

gdzie xx jest nazwą katalogu zawierającego przykładowe programy dla odpowiedniego łącznika, np. db2, od, dl itd.

Na serwerach AIX przykładowe programy dla pakietu narzędzi do tworzenia łączników są zorganizowane w następujący sposób:

```
/usr/lpp/cmb/samples/cpp/icm
/usr/lpp/cmb/samples/java/xx
/usr/lpp/cmb/samples/jsp/xx
/usr/lpp/cmb/samples/server/exi t
```

gdzie xx jest nazwą podkatalogu, np. beans , servlets itd.

Na serwerach Solaris przykładowe programy dla pakietu narzędzi do tworzenia łączników są zorganizowane w następujący sposób:

```
/opt/IBMcmb/samples/java/xx
/opt/IBMcmb/samples/jsp/xx
/opt/IBMcmb/samples/server/exi t
```

gdzie xx jest nazwą podkatalogu, np. beans , servlets itd.

| Przykładowe programy zawierają dokumentację opisującą programy oraz informacje o
| konfiguracji serwera (zmiennych środowiskowych, pamięci itd.), które są niezbędne przy
| pracy z przykładowym kodem.

Centrum informacyjne

Składnik Centrum informacyjne zawiera Centrum informacyjne Enterprise Information Portal. Jest to internetowa wersja biblioteki Enterprise Information Portal z możliwością przeszukiwania.

Rozdział 2. Klient administracyjny - wprowadzenie

Klient administracyjny zapewnia interfejs pomiędzy administracyjną bazą danych produktu EIP a administratorem EIP. W tej sekcji opisano wiele funkcji i cech klienta, pomocnych w administrowaniu programem EIP.

Niektóre z funkcji (np. definiowanie serwerów, zarządzanie użytkownikami) można wywołać za pomocą ikon umieszczonych w lewym panelu programu klienckiego. Dostęp do innych funkcji zapewnia pasek menu Narzędzia.

Pierwsze kroki w programie klient administracyjny systemu

Program *Pierwsze kroki* jest modulem dostarczającym w każdej instalacji programu EIP. Program *Pierwsze kroki* dostarcza przykładowych danych i wypełnia obiekty, nie ma więc potrzeby używania prawdziwych danych. Program *Pierwsze kroki* służy do eksplorowania definicji serwera, użytkowników i grup, a także innych funkcji. Dzięki temu łatwiej jest zrozumieć podstawową strukturę, wygląd i opcje klienta administracyjnego.

Administrowanie programem EIP

Administrator systemu może za pomocą klienta administracyjnego wykonywać wymienione poniżej czynności:

- definiowanie serwerów danych
- zarządzanie użytkownikami i grupami
- zarządzanie uprawnieniami i poziomami dostępu
- tworzenie szablonów wyszukiwania stowarzyszonego
- tworzenie jednostek stowarzyszonych
- tworzenie poddomen, jeśli domeny administracyjne są włączone
- obsługa przepływu pracy, jeśli jest włączony
- tworzenie obiektu tekstu stowarzyszonego w produkcie Content Manager w wersji 7.

Zarządzanie użytkownikami i grupami

Aby udostępnić użytkownikom dostęp do funkcji wyszukiwania dokumentów i pracy z nimi na wielu serwerach danych, należy utworzyć identyfikatory użytkowników i nadać im uprawnienia. Przypisując użytkownikom odpowiednie uprawnienia, administrator ma możliwość ograniczania dostępu użytkowników do danych przechowywanych w systemie.

Używanie narzędzi klienta administracyjnego

W tej sekcji opisano programy narzędziowe dostarczane wraz z klientem administracyjnym.

Konfiguracja LDAP

Po kliknięciu tej opcji, program EIP wyświetla okno, w którym dostępne są następujące karty:

- LDAP - za jej pomocą można zaimportować źródło danych z serwera LDAP oraz włączyć importowanie użytkowników i uprawnień z LDAP.
- Serwer - zawiera pola określające serwer LDAP, m.in. nazwę hosta, nazwę użytkownika, typ odwołania i inne.
- Uwierzytelnianie - zawiera pola służące do definiowania informacji SSL.

- Zaawansowane - definiuje ustawienia dla maksymalnej wielkości rekordu oraz limit czasu serwera.

Odwzorowanie użytkowników

Ta opcja pozwala na wyłączenie domyślnie włączonego odwzorowania użytkowników.

Edytor odwzorowania użytkowników stowarzyszonych

Edytor użytkowników stowarzyszonych wyświetla listę użytkowników i umożliwia odwzorowanie ich względem odpowiednich serwerów danych.

Przeglądarka szablonów wyszukiwania

Przeglądarka szablonów wyszukiwania dostarcza szczegółowych informacji na temat szablonów wyszukiwania. Przeglądarka dostarcza trzech opcji do przeglądania szczegółów szablonu wyszukiwania:

- Skojarzone odwzorowania (opcja domyślna) - dostarcza szczegółów na temat jednostek stowarzyszonych i innych informacji na temat szablonu wyszukiwania.
- Szablon wyszukiwania - zawiera szczegóły na temat operatora domyślnego, wartości domyślnych itd.
- Wyświetlane wyniki - zawiera informacje na temat wyświetlanej nazwy, szerokości wyświetlania, kolejności kryteriów itd.

Przeglądarka spisu zasobów

Wyświetla spisywanie zasobów dla wybranego serwera lub serwerów.

Przeglądarka protokołów

Aby wyświetlić protokół wygenerowany po odświeżeniu spisu zasobów serwera, należy użyć okna Przeglądarka protokołu. Protokół zawiera listę komunikatów wygenerowanych podczas znajdowania różnic między nowym a poprzednim spisem zasobów.

Usługi Aby włączyć eksplorację informacji i/lub przepływ pracy, należy wybrać Usługi.

Domeny administracyjne

Wybierz domeny administracyjne, aby je włączyć. Włączonych domen administracyjnych nie można wyłączyć.

Edytor typów MIME

Edytor typów MIME dla każdego serwera danych wyświetla listę następujących informacji:

- Klasa zawartości
- Rozszerzenie nazwy pliku
- Kolumna relacyjnej bazy danych (RDB)
- Typ MIME

Nazwy serwerów danych, wyświetlanych w edytorze typów MIME, są skracane i odpowiadają nazwom serwerów, które pojawiają się podczas definiowania nowego serwera. **Wskazówka:** DL jest skrótem nazwy serwera danych Content Manager wersja 7.1. V4 jest skrótem serwera danych Content Manager for AS/400.

Wartości wyświetlane w edytorze typów MIME można zmieniać, usuwać i dodawać nowe.

Edytor skojarzeń MIME dla aplikacji

Edytor skojarzeń MIME dla aplikacji może być używany do zmiany, dodawania i usuwania pięciu domyślnych skojarzeń typów MIME z aplikacjami. Wartości i ustawienia, zdefiniowane w tym edytorze, mają wpływ na przeglądarkę używaną przez aplikacje użytkowników końcowych.

Definicje typu serwera

Użyj tego narzędzia, aby zdefiniować własne serwery danych, opracowane przez własnych programistów systemowych.

Zmień ID/hasło DB2

Opcja ta służy do zmiany identyfikatora i hasła użytkownika DB2 tylko dla celów połączeń. Jest to inny identyfikator użytkownika niż identyfikator administratora.

Wprowadzenie do uprawnień

W tej sekcji omówiono uprawnienia w Informacyjnym Portalu Przedsiębiorstwa. Aby uzyskać dostęp do tych czterech uprawnień, należy rozwinąć ikonę Autoryzacja.

Wskazówka: ponieważ można administrować serwerem Content Manager w wersji 8 i portalem EIP w wersji 8 z tego samego klienta, wyświetlane są wszystkie uprawnienia po obu stronach klienta.

Uprawnienia

Wraz z klientem administracyjnym są dostarczane predefiniowane uprawnienia, grupy i zbiory uprawnień. Uprawnienia pozwalają użytkownikom i administratorom na uzyskanie określonego dostępu do określonych obiektów.

Uprawnienia

EIP udostępnia różne uprawnienia. Uprawnienie to prawo dostępu do określonego obiektu w określony sposób. Na przykład użytkownikom końcowym można przydzielić uprawnienia ItemAdd i ItemDelete, dzięki czemu będą oni mogli dodawać lub usuwać elementy z serwera danych. Aby wyświetlić uprawnienia, kliknij ikonę Uwierzytelnianie i kliknij dwukrotnie opcję Uprawnienia. Aby utworzyć uprawnienia, kliknij opcję Uprawnienia prawym przyciskiem myszy i kliknij opcję Nowe.

Grupy uprawnień

EIP udostępnia domyślne grupy uprawnień. Grupa uprawnień jest kolekcją powiązanych uprawnień. Przykładowo grupa uprawnień o nazwie Administer EIP zawiera pięć typowych uprawnień związanych z administrowaniem systemem EIP:

- EIPAdminServer
- EIPAdminEntity
- EIPAdminTextEntity
- EIPAdminTemplate
- EIPAdminInfoMining

Aby wyświetlić lub zmodyfikować uprawnienia przypisane wstępnie do grupy uprawnień, rozwiń ikonę uwierzytelnianie i kliknij dwukrotnie nazwę grupy uprawnień. Aby utworzyć grupy uprawnień, kliknij prawym przyciskiem myszy opcję Grupy uprawnień, a następnie opcję Nowa.

Zbiory uprawnień

Program EIP dostarcza wiele domyślnych zbiorów uprawnień. Zbiory uprawnień są kolekcjami uprawnień, które definiują role użytkowników. Na przykład zbiór uprawnień o nazwie ClientUserCreateAndDelete zawiera 17 uprawnień związanych z rolami użytkowników końcowych, takimi jak Delete (usuwanie elementu), ItemAdd (dodawanie elementu) itd. Po utworzeniu identyfikatora użytkownika klienta i przypisaniu mu zbioru uprawnień ClientUserCreateAndDelete użytkownik może zalogować się na serwerze danych i wykonywać jedną z 17 ról użytkowników, które znajdują się w zbiorze uprawnień. Aby wyświetlić lub zmodyfikować zbiory uprawnień, rozwiń ikonę Uwierzytelnianie, kliknij opcję Zbiory uprawnień i dwukrotnie kliknij nazwę zbioru uprawnień. Aby utworzyć zbiory uprawnień, kliknij prawym przyciskiem myszy opcję Zbiory uprawnień, a następnie opcję Nowa.

Administrowanie użytkownikami i grupami polega na przypisywaniu zbioru uprawnień do użytkownika lub grupy użytkowników. Po przypisaniu zbioru uprawnień do grupy użytkowników wszyscy jej członkowie będą mieli przypisany ten zbiór i będą mogli wykonywać znajdujące się w nim role.

Przełączanie widoku produktu i baz danych

Jeśli produkty Content Manager i Informacyjny Portal Przedsiębiorstwa są częścią rozwiązania w przedsiębiorstwie, obaj klienci administracyjni są dostępni z jednego interfejsu użytkownika. W przeszłości w takiej sytuacji należało uruchamiać dwóch niezależnych klientów. Przełączanie z widoku jednego klienta do drugiego jest wygodną metodą modyfikowania informacji, która ma zastosowanie do obydwu klientów.

Aby przełączyć z administrowania produktem EIP na produkt Content Manager bez wylogowania, należy w głównym oknie administracji systemem użyć menu rozwijanego umieszczonego nad lewym panelem i wybrać opcję Content Manager.

Aby przełączyć pomiędzy stowarzyszonymi bazami danych, należy w oknie klienta przejść do prawego panelu i dwukrotnie kliknąć ikonę stowarzyszonej bazy danych.

Można także administrować wieloma bazami danych bez wychodzenia z klienta i wpisywania się do innej bazy. Klient administracyjny wyświetla ikonę dla wszystkich administracyjnych baz danych znajdujących się w pliku cmbds.ini. Aby przełączyć się pomiędzy bazami danych, wystarczy kliknąć odpowiednią ikonę. Jeśli nowa baza danych wymaga innego identyfikatora użytkownika niż zalogowany w kliencie, zostanie wyświetlone okno z żądaniem wpisania innego identyfikatora.

Udoskonalenia w kliencie administracyjnym

W kliencie administracyjnym EIP wersja 8.2 wprowadzono szereg zmian, takich jak:

Poprawiono kreatorów i okna dialogowe

Dzięki nowym oknom dialogowym administrowanie użytkownikami jest łatwiejsze. Obsługa nowych kreatorów, którzy służą do definiowania i zmieniania jednostek stowarzyszonych i szablonów wyszukiwania, jest łatwiejsza. Użytkownicy mogą też nadal używać okien dialogowych obsługiwanych przez produkt EIP w wersji 7.1.

Współużytkowany klient administracyjny

Gdy na tym samym komputerze zainstalowany jest program EIP 8.2 i Content Manager w wersji 8.2, obydwa produkty mogą używać tych samych klientów administracyjnych i baz danych. Jeżeli tymi produktami zarządza jedna osoba, to do pracy wymagane jest tylko jedno zalogowanie się, po którym można przełączać klienta między dwoma aplikacjami. Można także przełączać się pomiędzy administracyjnymi bazami danych bez potrzeby wylogowania się z powtórnego zalogowania.

Administratorzy domen

Można utworzyć administratorów domen, którzy będą mieli uprawnienia w zdefiniowanej domenie.

Pojedyncze logowanie i obsługa LDAP

Program EIP używa obecnie usług Windows Active Directory lub serwera LDAP w celu umożliwienia użytkownikom uzyskania dostępu do wielu serwerów danych, nie wymagając wielokrotnego logowania się.

Łączenie klienta administracyjnego z lokalną administracyjną bazą danych

Po zainstalowaniu administracyjnej bazy danych na tym samym serwerze, na którym został zainstalowany klient administracyjny, informacje wymagane do podłączenia lokalnego klienta i serwera są już zapisane w pliku `cmbds.ini`, który służy do przechowywania informacji o połączeniu z bazą danych. Po zakończeniu instalacji nie ma potrzeby wykonania konfiguracji i można natychmiast nawiązać połączenie, wykonując informacje podane w niniejszej sekcji. **Wymaganie:** Po utworzeniu dodatkowych lokalnych baz danych przy użyciu programu instalacyjnego bazy danych EIP przed nawiązaniem połączenia z nową bazą danych należy samodzielnie zmodyfikować plik `cmbds.ini`, wprowadzając do niego wymagane informacje.

1. Kliknij kolejno opcje **Start-->Programy-->Enterprise Information Portal for Multiplatforms 8.2-->Administration**
2. Z rozwijanej listy Server wybierz lokalną bazę danych.
3. Wpisz ID administratora i hasło i kliknij przycisk OK.
4. Zostanie uruchomiony klient administracyjny systemu. **Wskazówka:** Jeśli była wykorzystywana aplikacja EIP First Steps, w lewym panelu klienta zostaną wyświetlone przykładowe bazy danych.

Łączenie klienta administracyjnego ze zdalną administracyjną bazą danych

Klienta administracyjnego EIP można połączyć ze zdalną bazą danych w systemie AIX, Windows lub Solaris na dwa sposoby:

- Za pośrednictwem serwera RMI (więcej informacji na ten temat zawiera publikacja *Planning and Installing Information Integrator for Content*).
- Definiując połączenie poprzez skatalogowanie bazy danych przy użyciu programu DB2 Configuration Assistant, a następnie definiując parametry połączenia z serwerem za pomocą programu narzędziowego EIP Server Configuration. Program ten kopiuje informacje, takie jak nazwa schematu bazy danych, nazwa aliasu, system operacyjny itd. do pliku o nazwie `cmbds.ini`. Po uruchomieniu klienta administracyjnego systemu z pliku `cmbds.ini` jest pobierana lista serwerów, do których można się zalogować.

Wymaganie: Każdą ze zdalnych baz danych należy skatalogować osobno. Każda zdalna baza danych musi być wymieniona w pliku `cmbds.ini`, zanim będzie można zawiązać z nią połączenie przy użyciu klienta administracyjnego.

Wskazówka: Doświadczeni użytkownicy mogą pominąć kroki wykonywane w programie narzędziowym Server Configuration Utility i mogą zmodyfikować plik `cmbds.ini` przy użyciu edytora tekstu. Plik `cmbds.ini` jest standardowo umieszczony w katalogu `C:\Program Files\IBM\CMgmt`.

Ważne: Jeśli osoba, która zainstalowała produkt, już skonfigurowała wartości katalogu bazy danych dla zdalnej bazy, z którą chcesz się połączyć, nie wykonuj dla tej bazy kroków DB2 CCA. Ale jeśli instalator nie wpisał wartości katalogu bazy danych lub jeśli chcesz się połączyć z dodatkową zdalną bazą danych, musisz użyć programu DB2CCA i zmodyfikować plik `cmbds.ini` podając parametry połączenia dla tej dodatkowej bazy danych.

Krok 1 - skataloguj zdalną bazę danych przy użyciu programu DB2 Configuration Assistant

Program DB2 Configuration Assistant (CCA) kataloguje zdalną bazę danych EIP w DB2. Aby wykonać tę operację, musisz znać nazwę zdalnego serwera, bazy danych i instancji bazy danych, oraz musisz zdefiniować alias dla zdalnej bazy danych.

Kroki od 1a do 1f wyjaśniają sposób lokalizowania nazwy bazy danych, nazwy schematu oraz numeru portu. Użytkownik musi znać nazwy i numery portów niezbędne do skonfigurowania połączenia między klientem administracyjnym a zdalną bazą danych.

1. Odszukaj informacje dotyczące połączenia ze zdalną bazą danych.
 - a. Zaloguj się do zdalnego serwera AIX, Windows lub Solaris używając identyfikatora użytkownika, który ma uprawnienia administratora DB2.
 - b. Wpisz `db2 list db directory`
 - c. Wybierz nazwę administracyjnej bazy danych, z którą chcesz nawiązać połączenie. Zapisz instancję db2, w której jest zainstalowana baza danych, ponieważ różnym instancjom odpowiadają różne numery portów.
 - d. Wpisz `db2 connect to <database> user <userID> using <password>`
 - e. Wpisz `db2 list tables` i zapisz nazwę schematu bazy danych (wymagane przez program narzędziowy dla konfiguracji serwera).
 - f. Sprawdź numer portu powiązany ze zdalną administracyjną bazą danych:

Dla systemu Windows:
 - 1) Otwórz Centrum sterowania DB2 Control Center na zdalnym serwerze z systemem Windows.
 - 2) Kliknij prawym przyciskiem myszy na instancjach dostępnych na lokalnej maszynie.
 - 3) Wybierz opcję konfiguracji komunikacji. .
 - 4) Naciśnij przycisk Właściwości, który znajduje się na prawo od opcji TCP/IP. W oknie zostanie wyświetlony numer portu.
Dla systemów AIX lub Solaris
 - 1) Wpisz `cd /usr/etc`
 - 2) Wpisz `cat services`
 - 3) Przewiń listę usług aż znajdziesz numer portu instancji zdalnej bazy danych. Na przykład, jeśli baza danych jest zainstalowana na instancji db2inst1 , port może mieć numer 50000.
 - 4)
2. Użyj programu DB2 Configuration Assistant, aby skatalogować zdalną bazę danych. Więcej informacji na ten temat zawierają pliki pomocy programu DB2CCA.
 - a. Zaloguj się do serwera Windows, na którym jest zainstalowany klient administracyjny. Do zalogowania się musisz użyć identyfikatora użytkownika, który ma pełne uprawnienia DB2ADM.
 - b. Uruchom program DB2 Configuration Assistant z menu Start-->Programy.
 - c. Postępuj zgodnie z instrukcjami programu DB2 Configuration Assistant, aby skatalogować i przetestować połączenie ze zdalną bazą danych.
 - d. Jeśli test połączenia DB2 CCA zostanie pomyślnie przeprowadzony, wykonaj kroki opisane w sekcji “Krok 2 - użyj programu Server Configuration Utility”, lub zmodyfikuj bezpośrednio plik cmbds.ini, aby zdefiniować parametry połączenia ze zdalną bazą danych przechowywane w pliku cmbds.ini

Krok 2 - użyj programu Server Configuration Utility

Program narzędziowy do konfiguracji serwera prosi o podanie informacji dotyczących połączenia (nazwa hosta, numer portu itp.) ze zdalną bazą danych i zapisuje te dane w pliku cmbds.ini .

1. Kliknij kolejno opcje **Start-->Programy-->IBM Enterprise Information Portal for Multiplatforms-->Server Configuration Utility**.

2. Wpisz w polach wymagane informacje (patrz Tabela 4).

Tabela 4. Narzędzie do konfiguracji serwera

Pole	Informacja	Notes
Serwer	Wybierz typ bazy danych: Content Manager lub EIP.	Serwer oznacza typ bazy danych, a nie nazwę serwera, na którym jest zainstalowana baza danych. Wskazówka: Klienta administracyjnego można używać do administrowania obiema bazami danych tylko wtedy, gdy klienci administracyjni programu Content Manager i EIP są zainstalowani na tym samym komputerze.
Nazwa serwera	Wpisz nazwę aliasu bazy danych, z którą chcesz się połączyć. Wymaganie: Musisz użyć tej samej nazwy aliasu zdefiniowanej w DB2CCA.	Alias tworzy unikalną nazwę, która identyfikuje zdalną bazę danych na Twojej stacji roboczej. Nazwy aliasów nie mogą być dłuższe, niż 8 znaków. Na przykład, jeśli nazwa zdalnej bazy danych to ICMNLSDB, alias może mieć nazwę REMOTE1.
Nazwa schematu	Wpisz nazwę schematu przypisaną podczas tworzenia zdalnej bazy danych.	Dla baz danych EIP i Content Manager domyślną nazwą schematu jest ICMADMIN.
Nazwa hosta	Wpisz nazwę komputera, na którym została zainstalowana zdalna baza danych.	Wpisz pełną nazwę hosta lub adres IP komputera, na którym została zainstalowana zdalna baza danych.
System operacyjny	Wybierz z rozwijanej listy system operacyjny.	Wybierz opcję AIX, Sun Solaris lub Windows. System EIP 8.2 nie obsługuje systemu OS/390.
Numer portu	Wpisz numer portu skojarzony ze zdalną bazą danych.	Dla baz danych EIP i Content Manager zainstalowanych w systemach Windows, AIX i Solaris standardowym numerem portu jest 50000.
Nazwa zdalnej bazy danych	Wpisz nazwę zdalnej bazy danych. Można używać tylko wielkich liter.	Dla baz danych EIP i Content Manager domyślną nazwą bazy danych jest ICMNLSDB.
Nazwa węzła	Wpisz nazwę zdalnej bazy danych programu EIP lub Content Manager.	Nazwa węzła jest unikalną nazwą przypisaną zdalnej bazie danych, podobną do aliasu utworzonego dla zdalnej bazy danych. Aby sprawdzić nazwę węzła dla bazy danych zainstalowanej na serwerze z systemem operacyjnym Windows, AIX lub Solaris, wykonaj następujące czynności: a. Otwórz sesję wiersza komend db2. b. Przy znaku zgłoszenia db2=> wpisz LIST NODE DIRECTORY c. Baza DB2 wyświetli nazwy węzłów i inne dane dotyczące wszystkich baz danych zainstalowanych lub zdefiniowanych na zdalnym serwerze.

Tabela 4. Narzędzie do konfiguracji serwera (kontynuacja)

Pole	Informacja	Notes
Włącz jednokrotne logowanie się	Zaznacz tę opcję, jeśli podczas instalacji zostało aktywowane jednokrotne logowanie się.	Domyślnie opcja ta nie jest zaznaczona.
Opcje ochrony	Włącz uwierzytelnianie klienta, jeśli podczas tworzenia bazy danych ta opcja została aktywowana.	Domyślne ustawienie to Serwer.

3. Kliknij przycisk OK.

Krok 3 - testowanie połączenia ze zdalną bazą danych

1. Zaloguj się do serwera Windows, na którym jest zainstalowany klient administracyjny.
2. Kliknij kolejno opcje **Start-->Programy-->Enterprise Information Portal for Multiplatforms 8.2-->Administration**.
3. Z rozwijanej listy Server wybierz alias zdalnej bazy danych. Nazwa odpowiada aliasowi zdefiniowanemu w programie narzędziowym do konfiguracji serwera i w programie DB2 Configuration Assistant.
4. Wpisz ID użytkownika i hasło zdefiniowane dla zdalnej bazy danych.
5. Kliknij przycisk OK. Zostanie otwarty klient administracyjny.

Definiowanie typów dokumentów

Program EIP zapewnia obsługę przeglądania niektórych typów dokumentów. Dokumenty, których typ został zdefiniowany na serwerze, można otwierać w aplikacjach, w których zostały utworzone. Na przykład, jeśli dokumenty Lotus Word Pro są przechowywane na serwerze Content Manager OnDemand, można tak skonfigurować program EIP, aby dokumenty z rozszerzeniem .lwp były otwierane w programie Lotus Word Pro zamiast w przeglądarce dokumentów klienta.

Aby zdefiniować typ dokumentu, zmodyfikuj plik cmbcc2mime.ini, znajdujący się w katalogu CMBROOT, zazwyczaj x:\Program Files\IBM\CMgmt. Plik ten zawiera instrukcje dotyczące tworzenia własnych definicji MIME. Umożliwia przekształcenie klasy zawartości na strumień typu MIME, co umożliwia odczytywanie w programie klienckim zawartości serwerów danych.

Ważne: Podczas uruchamiania aplikacji na podstawie typu MIME wyświetlany jest tylko obiekt podstawowy. Nie są wyświetlane żadne znaczniki występujące w dokumencie. Jeśli dokument składa się z wielu części, wyświetlana jest tylko pierwsza część. Typy MIME w obu plikach muszą być takie same.

Modyfikowanie pliku typu MIME (cmbcc2mime.ini)

Dodając typy MIME serwera, należy sprawdzić, czy dodawany typ dokumentu jest typem MIME utworzonym dla tego pliku. Więcej informacji na ten temat zawiera serwis WWW <ftp://ftp.isi.edu/in-notes/iana/assignments/media-types>.

Aby dodać nowe wartości w pliku `cmbcc2mime.ini`, należy wykonać następujące czynności:

1. W edytorze tekstu otwórz plik `cmbcc2mime.ini`.
2. Dla wartości definiowanych przez użytkownika należy używać następującego formatu:
 - Klasa zawartości rozpoczyna się od wartości 4096.
 - Znak równości (=) znajduje się po wartości klasy zawartości.
 - Po znaku równości należy podać typ MIME. Jeśli nie jest to standardowy typ MIME dla tej klasy zawartości, należy wykonać następujące czynności:
 - a. Typ MIME składa się z typu i podtypu. Prawidłowe typy to: application, text, image, model, message, audio i video.
 - b. Definicję typu kończy znak ukośnika (/).
 - c. Aby utworzyć podtyp, należy symbol określający dokument poprzedzić symbolem (x-), na przykład:
`x-mydocumentclass (4096=application/x-mydocumentclass)`
 - Powtórz czynności 2b i 2c dla każdego nowego typu MIME.

Wskazówka: Serwery danych OnDemand odwzorowują rozszerzenia plików, a nie wartości liczbowe klas danych, do strumieni typów MIME.

Rozdział 3. Używanie narzędzi klienta administracyjnego

W tej sekcji zostało przedstawione wyjaśnienie kilku podstawowych czynności wykonywanych przez administratorów EIP.

Tworzenie wyszukiwania stowarzyszonego

Wyszukiwanie stowarzyszone jest żądaniem wydanym przez aplikację kliencką, które uruchamia wyszukiwanie jednocześnie na kilku serwerach danych. Program EIP udostępnia narzędzia do tworzenia szablonów wyszukiwania dla wyszukiwania stowarzyszonego. Ponieważ każdy serwer danych przechowuje i organizuje informacje w inny sposób, szablon wyszukiwania musi uwzględniać te różnice dla każdego serwera. Szablon wyszukiwania, przeszukując serwery danych, odwzorowuje jednostki stowarzyszone i ich atrybuty stowarzyszone na atrybuty rodzime.

Tworzenie wyszukiwań stowarzyszonych polega na wykonaniu następujących czynności:

- Definiowanie połączeń z serwerami danych przy użyciu łączników EIP.
- Tworzenie jednostek stowarzyszonych.
 - Definiowanie jednostek stowarzyszonych
 - Tworzenie atrybutów stowarzyszonych
 - Odwzorowywanie atrybutów stowarzyszonych na atrybuty rodzime
 - Przypisywanie parametrów
- Tworzenie szablonów wyszukiwania
 - Definiowanie szablonu wyszukiwania
 - Definiowanie kryteriów wyszukiwania
 - Definiowanie ustawień szablonu
 - Przypisywanie dostępu do użytkowników

EIP wersja 8.2 zawiera dwa kreatory, które znacznie ułatwiają tworzenie jednostek stowarzyszonych i szablonów wyszukiwania w porównaniu do wcześniejszych wersji EIP. Kreator jednostki stowarzyszonej zawiera spis zasobów serwera, który można filtrować w celu ułatwienia znalezienia atrybutów rodzimych. Generuje także poprawne parametry dla atrybutów stowarzyszonych, zmniejszając możliwość ich błędnego skonfigurowania. Kreator szablonów wyszukiwania umożliwia definiowanie kryteriów wyszukiwania. Ułatwia także określenie sposobu wyświetlania i funkcjonowania kryteriów wyszukiwania. Umożliwia nawet wyświetlenie szablonu wyszukiwania w takiej postaci, jak może wyglądać w aplikacji klienckiej. Dostępne są także okna dialogowe tworzenia jednostek stowarzyszonych i szablonów wyszukiwania z wersji 7.1 produktu EIP.

Wszystkie kreatory, okna dialogowe oraz pola są opisane w pomocy elektronicznej dla EIP.

Definiowanie serwerów

Aby było możliwe połączenie się z serwerem i wykonanie spisu zasobów serwera, należy zdefiniować serwer. Kliknięcie prawym przyciskiem myszy ikony Serwer, a następnie opcji Nowy, powoduje wyświetlenie w kliencie wszystkich łączników obsługiwanych przez portal EIP. Przed zdefiniowaniem serwera należy poznać odpowiedzi na następujące pytania:

- Które łączniki zostały wybrane przez program instalacyjny? Lista zainstalowanych łączników znajduje się w pliku konfiguracyjnym cmbcs.ini. W serwerze Windows

domyślną ścieżką do tego pliku jest `x:\Program Files\IBM\CMgmt`. Informację o położeniu plików `cmbscs.ini` w systemach AIX i Solaris można uzyskać od ich administratorów.

- Czy program instalacyjny wybrał łączniki zdalne czy lokalne? Plik `cmbscs.ini` zawiera lokalne i zdalne typy łączników.
- Jeżeli system skonfigurowany jest dla RMI, czy serwer RMI jest uruchomiony? Aby uruchomić RMI na lokalnym serwerze RMI, wybierz kolejno opcje **Start**→**Programy**→**IBM Enterprise Information Portal for Multiplatforms 8.2**→**Start RMI servers**. Jeśli system używa zdalnych łączników RMI należy w pliku `cmbsvclient.ini` znaleźć serwer, na którym są one zainstalowane. Więcej informacji na ten temat można uzyskać od administratora serwera RMI.
- Jeśli osoba instalująca portal EIP zainstalowała łącznik CM for AS/400, jakie informacje znalazły się w tabeli sieciowej o nazwie `frnolint.tbl`? Plik `frnolint.tbl` dla systemu AS/400 znajduje się w katalogu określonym przez zmienną `%CMBROOT%`.
- Definiując serwery danych zawierające relacyjne bazy danych, takie jak Content Manager wersja 8 i DB2, DataJoiner i Information Catalog, należy skatalogować lub dodać bazę danych ze stacji roboczej, na której używany jest klient.

Poniżej umieszczono ogólną listę czynności, które należy wykonać podczas definiowania serwera:

1. Kliknij prawym przyciskiem myszy serwer i wybierz **Nowy**.
2. Z wyświetlonej listy wybierz serwer. Wyświetlone zostanie okno Nowy serwer.
3. Na karcie Ogólne w polu Nazwa serwera wpisz nazwę i opis serwera. Dla niektórych serwerów należy podać tylko nazwę bazy danych. Dla pozostałych należy podać pełną nazwę serwera, na którym została zainstalowana baza danych.
4. Podaj parametry inicjujące, jeżeli jest to wymagane. Niektóre serwery wymagają podania parametrów inicjujących, np. łańcuchów tekstowych połączenia i konfiguracji. Pozostałe serwery wymagają jedynie podania nazwy bazy danych.
5. Kliknij opcję testowania połączenia z serwerem. Program EIP zaloguje się do niektórych serwerów, używając identyfikatora użytkownika i hasła podanego podczas uruchamiania klienta administracyjnego. Jeżeli do połączenia wymagane jest podanie innych danych, program EIP poprosi o wpisanie poprawnego identyfikatora użytkownika i hasła dla definiowanego serwera danych.

Wskazówka: Można także zdefiniować nowy typ serwera, który nie został wstępnie zdefiniowany, ale wymaga to dostarczenia klas połączeniowych języka Java lub C++ i klasy definicji serwera dla nowego typu serwera. Wymagany jest także łącznik Java, służący do uruchamiania spisu zasobów serwera. Instrukcje dodawania serwerów danych można znaleźć w podręcznikach *Workstation Application Programming Guide* i elektroniczny skorowidz interfejsu API.

Jeżeli konfiguracja serwera danych nie powiodła się, należy zapoznać się z publikacją *Komunikaty i kody*, zawierającą więcej informacji na temat rozwiązywania problemów i komunikatów o błędach.

Można również poprosić o pomoc administratora serwera, z którym ma być nawiązane połączenie.

Wytyczne do definiowania serwerów

W tej sekcji umieszczono wskazówki pomocne podczas początkowego definiowania serwera.

Łączenie z relacyjną bazą danych DB2

Sekcja ta dotyczy serwerów DB2, DataJoiner, JDBC, ODBC, Information Catalog, i Content Manager wersja 7 i 8.

- **Ważne:** Przed zdefiniowaniem serwera należy skatalogować każdą bazę danych DB2. Do tego celu można użyć programu DB2 CCA albo wiersza komend DB2. Więcej informacji na ten temat może podać administrator DB2.
- W polu Nazwa serwera (na karcie Ogólne) należy wpisać nazwę bazy danych, z którą ma być ustanowione połączenie. Do wpisywania nazwy serwera należy używać wielkich liter.
- Po zdefiniowaniu DB2, DataJoiner, JDBC, ODBC i Information Catalog, kliknij kartę Parametry inicjujące i wpisz nazwę schematu, który jest skojarzony z tabelami połączonej bazy danych, np. SCHEMA=ICMADMIN.
- Podczas definiowania serwera Content Manager wersja 7.1 lub wersja 8.2 wymagane jest tylko wpisanie nazwy bazy danych. Nie zmieniaj ustawień domyślnych na karcie parametrów inicjujących.
- W przypadku definiowania serwera Content Manager w wersji 7.1 na dysku lokalnym w katalogu x:\CMBROOT musi znajdować się tabela sieciowa o nazwie frnlint.tbl. Tabela sieciowa zawiera nazwę hosta, numer portu i informacje o typie serwera, których portal EIP potrzebuje do zlokalizowania serwera bibliotecznego Content Manager w wersji 7.1 i załogowania się do niego. W przypadku definiowania wielu serwerów Content Manager w wersji 7.1, dla każdego z nich musi być osobna pozycja w pliku frnlint.tbl.
- Aby połączyć się z bazą danych DB2 DataJoiner, należy upewnić się, że metoda uwierzytelniania w portalu Enterprise Information Portal jest zdefiniowana jako serwer instancji bazy danych zdefiniowany w bazie danych DB2 Universal Database.
- Aby połączyć się z DataJoiner 2.1, należy pobrać program bind (ze strony WWW programu DataJoiner) i powiązać bazę danych DataJoiner przed definiowaniem serwera.

Łączenie z serwerem wyszukiwania tekstowego

Aby zdefiniować serwer wyszukiwania tekstowego, należy najpierw zdefiniować serwer programu Content Manager wersja 7.1, który jest skojarzony z serwerem wyszukiwania tekstowego.

Nazwę serwera wyszukiwania tekstowego wpisuje się w polu listy rozwijanej "Wybierz nazwę powiązanego serwera Content Manager wersja 7.1". Pole to znajduje się w zakładce Serwer powiązany.

Aby portal EIP mógł połączyć się z serwerem Content Manager w wersji 7.1 i serwerem wyszukiwania tekstowego, muszą one być zainstalowane i uruchomione.

Łączenie z wieloma serwerami Content Manager dla AS/400

Jeżeli używanych jest kilka serwerów AS/400, to dodatkowe serwery muszą być wpisane do tabeli sieciowej. Tabela sieciowa (frnlint.tbl) znajduje się w katalogu x:\<cmbroot>. Dla nowego serwera wpisz jego nazwę, typ połączenia (np TCP/IP), nazwę hosta, port oraz typ. Instalując pierwszy serwera należy wpisać wartości określające serwer, nazwę hosta i port, umożliwiając utworzenie pliku frnlint.tbl.

Poniżej znajduje się typowy przykład informacji przechowywanych w pliku frnlint.tbl:

```
/* Tabela sieciowa VI/400 */
SERVER: VI400 REMOTE TCPIP
        HOSTNAME = vi400
        PORT      = 29000
        SERVER_TYPE = FRNLS400
```

Konfigurowanie łącznika wyszukiwania rozszerzonego

Informacje, które należy podać w celu zdefiniowania serwera wyszukiwania rozszerzonego, zależą od dwóch czynników:

- typu serwera WWW, na którym zainstalowano serwer wyszukiwania rozszerzonego - Domino Web Server, WebSphere, IIS.
- numeru portu zdefiniowanego dla tego serwera WWW.

Podczas definiowania łącznika wyszukiwania rozszerzonego, należy wykonać następujące czynności:

1. W polu Nazwa serwera (na karcie Ogólne) należy wpisać pełną nazwę serwera WWW, na którym zainstalowano serwer wyszukiwania rozszerzonego.
2. Na karcie parametrów inicjujących, w polu Numer portu wpisz 80 (jeżeli taki port został określony podczas instalowania serwera wyszukiwania rozszerzonego).
3. W polu ID aplikacji wpisz **Demo**. Wpisz nazwę, jak w przykładzie.
4. W polu Hasło wpisz **Demo**.
5. W polu dodatkowych parametrów:
 - a. Nie zmieniaj dwóch średników, jeżeli wiesz, że serwer wyszukiwania rozszerzonego jest zainstalowany na serwerze Domino Web Server i że program instalacyjny użył domyślnych ustawień portu dla serwera WWW wyszukiwania rozszerzonego.
 - b. Poniżej umieszczono informacje opisujące sposób modyfikowania pola parametrów dodatkowych serwerów wyszukiwania rozszerzonego, zainstalowanych przy użyciu ustawień użytkownika.

Możesz skonfigurować łącznik wyszukiwania rozszerzonego razem z serwerem aplikacji WebSphere. Jeśli wyszukiwanie rozszerzone działa używając innego portu, niż 6001, albo jeśli nazwa serwera wyszukiwania rozszerzonego jest różna od nazwy serwera WWW, należy skonfigurować łącznik w taki sposób, aby używał poprawnych ścieżek względnych dla serwletu ES oraz właściwego numeru portu i nazwy serwera ES.

Możesz skonfigurować łącznik wyszukiwania rozszerzonego razem z serwerem aplikacji WebSphere. Jeśli wyszukiwanie rozszerzone działa używając innego portu, niż 6001, albo jeśli nazwa serwera wyszukiwania rozszerzonego jest różna od nazwy serwera WWW, możesz wykonać przedstawione poniżej pliki, aby utworzyć plik konfiguracyjny, np. `desclient.cfg`.

Przejdź do katalogu, w którym znajdują się aplikacje i przykłady. Utwórz plik konfiguracyjny, np. `desclient.cfg`. Plik ten nie jest dostarczany z programem Enterprise Information Portal.

W pliku `desclient.cfg` dodaj pozycję `DESHOSTNAME=(nazwa hosta ES)`

Jeśli wyszukiwanie rozszerzone jest skonfigurowane z serwerem aplikacji Domino, ustaw wartość `DESREQURI=/servlet/ESAdmin`. Jeśli wyszukiwanie rozszerzone jest skonfigurowane z serwerem aplikacji WebSphere, wpisz pozycję `DESREQURI=/lotuskms/ESAdmin` gdzie *servlet* jest ścieżką na serwerze HTTP, w której jest zainstalowany łącznik wyszukiwania rozszerzonego.

Jeżeli serwerem aplikacji jest WebSphere, zmienna `DESREQURI` powinna być ustawiona na `/lotuskms/ESAdmin` zamiast `/servlet/ESAdmin`.

Jeżeli źródła wyszukiwania rozszerzonego mają być używane z klientów uszczuplonych, zdefiniuj dodatkowy parametr:

`"DESCFGPATH=<bezwzględna ścieżka pliku desclient.cfg>"`

w oknach dialogowych definicji serwera DES w kliencie administracyjnym.

Jeżeli chcesz uruchomić przykłady wyszukiwania rozszerzonego, przekaż bezwzględną ścieżkę pliku `desclient.cfg` w argumentach wiersza komend.

Przykład 1:

```
TConnectDES es.stl.ibm.com użytkownik hasło
PORT=80;DESAPPID=Demo;DESAPPPW=hasło;DESCFGPATH=<bezwzględna ścieżka pliku desclient.cfg>;
```

Przykład 2:

```
java TConnectDES es.stl.ibm.com użytkownik hasło  
PORT=80;DESAPPID=Demo;DESAPPPW=hasło;DESCFGPATH=<bezwzględna ścieżka pliku desclient.cfg>;
```

Definiowanie serwera Information Catalog

Przed zdefiniowaniem serwera należy skatalogować serwer katalogu informacji. W polu Nazwa serwera wpisz, na przykład, PRZYKLAD1. W zakładce Parametry inicjowania wpisz SCHEMA=<nazwa schematu powiązanego z serwerem PRZYKLAD1>.

Definiowanie serwera OnDemand

Serwer OnDemand i demon serwera bibliotecznego muszą być uruchomione przed zdefiniowaniem serwera OnDemand. Aby zweryfikować, czy serwer i demon są uruchomione, można użyć komendy ping.

Na karcie Ogólne wpisz pełną nazwę serwera WWW, na którym zainstalowano serwer OnDemand.

W zakładce Parametry inicjowania wpisz numer portu przypisany podczas instalacji serwera OnDemand. Jeśli osoba instalująca serwer OnDemand wybrała domyślną wartość portu (0), w polu Numer portu wpisz 0. Jeśli osoba ta wpisała inny numer portu, wpisz ten numer, poprzedzając go znakiem #. Na przykład # 5000.

Jeżeli definiowany jest serwer, który był zainstalowany na komputerze AS/400 z systemem operacyjnym w wersji 4, trzeba w polu informacji dodatkowych wpisać dodatkowy parametr: STATECONNECT=#1 .

Jeżeli definiowany jest serwer, który był zainstalowany na komputerze z systemem operacyjnym OS/390 w wersji 2.1, wpisz numer portu wybrany podczas instalacji tego serwera OnDemand.

Serwer OnDemand wymaga, aby gniazdo było cały czas w stanie połączenia.

Definiowanie serwera wyszukiwania tekstowego

Aby zdefiniować serwer wyszukiwania tekstowego, należy najpierw zdefiniować serwer programu Content Manager wersja 7, który jest skojarzony z serwerem wyszukiwania tekstowego.

Na karcie Ogólne w polu Nazwa serwera wpisz nazwę serwera. Z pola listy rozwijanej w zakładce Serwer powiązany wybierz powiązany serwer Content Manager w wersji 7.1.

Aby portal EIP mógł połączyć się z serwerem Content Manager w wersji 7.1 i serwerem wyszukiwania tekstowego, muszą one być zainstalowane i uruchomione.

Definiowanie serwera Domino.Doc

W polu Nazwa serwera, wpisz ścieżkę do serwera i biblioteki na serwerze Domino.Doc. Na przykład: oakley/DominoDoc1/Lib.nsf.

Jeśli używane są łączniki lokalne, pakiet Domino Doc Desktop Enabler musi być zainstalowany na stacji roboczej z klientem EIP. Jeżeli używany jest serwer RMI, pakiet Domino Doc Desktop Enabler musi być zainstalowany na serwerze RMI. Pakiet Domino Doc Desktop Enabler musi być w tej samej wersji, co serwer Domino Doc.

Nie zmieniaj dwóch średników na karcie parametrów inicjujących.

Definiowanie serwera ImagePlus dla OS/390

Jeżeli definiowany jest serwer ImagePlus dla OS/390, w celu uzyskania połączenia z serwerem należy znać następujące parametry. Na poniższej liście umieszczono wartości przykładowe:

- Numer portu FAF: 3061
- Identyfikator aplikacji FAF: 01
- Protokół FAF: 4000
- Adres IP FAF: 9.67.43.83
- CICS menedżera dystrybucji obiektów: 4000
- Adres IP menedżera dystrybucji obiektów: 9.67.43.83
- Numer portu menedżera dystrybucji obiektów: 3082
- Identyfikator terminalu menedżera dystrybucji obiektów: *pozostaw to pole puste*
- Parametry dodatkowe: *FAFSITE=CS61;;*

Używanie śledzenia z programem Content Manager ImagePlus for OS/390

Jeśli nie można połączyć się z serwerem Content Manager ImagePlus for OS/390, w rozwiązaniu problemu może pomóc śledzenie. Jeśli zainstalowano łącznik dla Content Manager ImagePlus for OS/390, można włączyć śledzenie ImagePlus for OS/390, modyfikując plik `eypapi.ini` znajdujący się w katalogu określonym przez zmienną `cmbroot`.

Plik `eypapi.ini` zawiera następujące wiersze:

```
; ścieżka, w której umieszczone są pliki IPFAF
; (BEZ kończącego znaku '\')
; -- domyślnie jest to <katalog główny>\
;
IPFAFPath=d:\cmbroot
; Flagi protokołowania (pliki EYPLmdd.LOG)
; -- domyślnie protokołowanie jest wyłączone (0)
; -- 0 protokołowanie wyłączone
; -- 1 pliki protokołu tworzone są tylko dla błędów
; -- 2 pliki protokołu tworzone są dla wszystkich zdarzeń
;
Logging = 0

;-----
;
; Flagi protokołowania typów parametrów FAF utworzonych za pomocą API
; -- domyślnie protokołowanie jest wyłączone (0)
; -- 0 protokołowanie wyłączone
; -- 1 protokołowanie typów parametrów FAF
;
FafTypeLogs = 0
```

IPFAFPath

Określa katalog, w którym zapisywane są pliki protokołów. Pliki te są nazywane według schematu:

`EYPmdd.LOG`

gdzie *mmdd* jest miesiącem i dniem utworzenia pliku.

Logging

Określa, kiedy plik protokołu jest tworzony.

- 0** Protokołowanie wyłączone. Jest to wartość domyślna.
- 1** Tworzy pliki protokołów zawierające wyłącznie błędy.
- 2** Tworzy pliki protokołów zawierające wszystkie zdarzenia.

FafTypeLogs

Określa protokołowanie dla parametrów FAF, utworzonych za pomocą API.

- 0 Protokołowanie wyłączone. Jest to wartość domyślna.
- 1 Protokołowanie typów parametrów FAF.

Praca z łącznikiem OnDemand: strojenie TCP/IP i gniazdek

Przy łączeniu z serwerem OnDemand mogą pojawić się problemy z wydajnością spowodowane przez system Windows. Podczas powtarzających się operacji wyszukiwania i pobierania danych z serwera OnDemand są otwierane i zamykane gniazdko systemu Windows. Duży ruch między EIP a serwerem OnDemand może ulec ograniczeniu wskutek istnienia dwóch domyślnych ustawień systemu Windows:

- Gdy aplikacja zamyka gniazdko systemu Windows, system ten nadaje mu status TIME_WAIT na 240 sekund. W tym czasie gniazdko nie może zostać ponownie użyte.
- System Windows ogranicza liczbę portów dostępnych dla aplikacji do 5000.

Aby uniknąć związanych z tym problemów, należy za pomocą edytora rejestru Windows zmienić wartości określające limit czasu oczekiwania i liczbę portów.

- Zmień wartość czasu oczekiwania z 240 sekund na niższą (poprawny zakres to 30-300 sekund). Odpowiedni klucz rejestru to: HKEY_Local_Machine\System\CurrentControlSet\services\Tcpip\Parameters\TcpTimedWaitDelay .
- Zwiększ maksymalną liczbę portów z domyślnych 5000 na większą wartość (dopuszczalny zakres to 5000-65534). Odpowiedni klucz rejestru to: HKEY_Local_Machine\System\CurrentControlSet\services\Tcpip\Parameters\MaxUserPort

Więcej informacji na temat parametrów TcpTimedWaitDelay i MaxUserPort zawiera dokumentacja systemu Windows.

Praca z łącznikiem wyszukiwania rozszerzonego

Ta sekcja opisuje zmiany w łączniku wyszukiwania rozszerzonego w systemie EIP wersja 8.2.

Ustawienia narodowe są obsługiwane poprzez przekazanie odpowiedniej wartości w kluczu DESLOCALE. Przy bezpośrednim wywoływaniu łącznika tę parę wartości można przekazać w wierszu komend. Można ją ustawić w **dodatkowych** argumentach dla właściwości wyszukiwania rozszerzonego.

Wskazówka Oprogramowanie serwera wyszukiwania rozszerzonego znajduje się w pudełku produktu EIP wersja 8.2.

Tworzenie jednostek stowarzyszonych

Po zdefiniowaniu połączeń z serwerami danych kolejnym krokiem przy tworzeniu wyszukiwania stowarzyszonego jest utworzenie jednostki stowarzyszonej, która będzie stanowić bloki składowe szablonów wyszukiwania. W tej sekcji zostały opisane jednostki stowarzyszone i sposób ich tworzenia przy użyciu kreatora.

Informacje o jednostkach stowarzyszonych

Większość użytkowników aplikacji klienckich nie chce szukać potrzebnych im informacji osobno na każdym z serwerów. Zamiast tego chcą mieć możliwość uruchomienia pojedynczego wyszukiwania stowarzyszonego. Dzięki szablonom wyszukiwania użytkownicy mogą zapakować swoje wyszukiwanie do pojedynczego zapytania. Administrator EIP może tworzyć takie szablony wyszukiwania dla aplikacji klienckich. Przed utworzeniem szablonu

wyszukiwania należy najpierw utworzyć jednostki stowarzyszone, które odwzorowują swoje stowarzyszone atrybuty na atrybuty rodzime serwerów danych.

Na przykład DB2 przechowuje informacje w tabelach, których kolumny reprezentują atrybuty informacji przechowywanych w tabeli. Tabela o nazwie **Dane_klientów** może zawierać kolumny o nazwach **NrPolisy**, **Adres**, **Telefon** i **Zawód**.

Z drugiej strony program Content Manager zamiast tabel i kolumn posługuje się elementami i atrybutami. Te same informacje, które są przechowywane w DB2, mogą być przechowywane w obiekcie o nazwie **DaneKlt**. Jego atrybutami mogą być **NazwaKlt**, **Konto**, **AdresDomowy**, **TelDomowy** i **Zawód**. W obu tych przypadkach zapisywane są identyczne informacje, ale są one w różny sposób identyfikowane.

EIP rozwiązuje problem związany z tym, że różne serwery danych przechowują identyczne informacje w różny sposób. Do tego celu są używane jednostki stowarzyszone. Jednostka stowarzyszona nie przechowuje danych; przechowuje metadane dotyczące sposobu przechowywania danych przez każdy z serwerów. Po utworzeniu jednostki stowarzyszonej należy przypisać wszystkie jej atrybuty do odpowiadających im atrybutów rodzimych na serwerze danych, który ma odpowiadać na zapytanie.

Dla powyższego przykładu można utworzyć jednostkę stowarzyszoną o nazwie **Polisa**, z atrybutami stowarzyszonymi **Nazwa_Polisy**, **Numer_Polisy**, **Adres_Domowy** i **Zawód**. Następnie atrybuty stowarzyszone można przypisać do odpowiadających im atrybutów rodzimych.

EIP może wygenerować spis zasobów serwera zawierający te informacje. Kreator tworzenia jednostek stowarzyszonych umożliwia uzyskanie spisu zasobów serwera, który można filtrować w serwerach danych. Filtrowanie jest dostępne tylko wówczas, jeśli użyty został ten kreator. Nie będzie ono dostępne, jeśli jednostka stowarzyszona zostanie utworzona samodzielnie (tj. bez pomocy kreatora). Po wygenerowaniu spisu można rozpocząć tworzenie przypisań między atrybutami stowarzyszonymi a atrybutami rodzimymi.

Samo dokonanie przypisań nie wystarcza, ponieważ każdy z rodzimych atrybutów może mieć różne właściwości. Atrybuty mogą być następujące: (1) dopuszczalne wartości puste, (2) dopuszczalny w zapytaniach, (3) dopuszczalna aktualizacja i (4) dopuszczalne w wyszukiwaniu tekstowym. W zależności od typu wybranych danych mogą się również pojawić opcje dotyczące długości danych, ich precyzji, skali oraz wartości minimalnej i maksymalnej.

Po zdefiniowaniu tych właściwości nie można zmienić ich na bardziej restrykcyjne niż właściwości które są już przypisane do atrybutów stowarzyszonych. Kreator udostępnia domyślne właściwości, które spełniają te kryteria. Jeżeli po dostosowaniu właściwości dla domyślnego atrybutu stowarzyszonego niezbędne okaże się przywrócenie domyślnych właściwości sugerowanych przez kreatora, nadal można wybrać wartości domyślne.

Podsumowując, atrybuty stowarzyszone przypisuje się do odpowiadających im atrybutów rodzimych na wielu serwerach danych. Każdy z atrybutów stowarzyszonych zastępuje wszystkie właściwości atrybutów rodzimych. Po utworzeniu atrybutu stowarzyszonego powstaje ścieżka do informacji przechowywanych na różnych serwerach danych. Następnie można użyć jednostek stowarzyszonych do tworzenia szablonów wyszukiwania dla określonych zapytań.

Korzystanie z kreatora tworzenia jednostek stowarzyszonych

Kreator jednostek stowarzyszonych jest nowością wprowadzoną w produkcie EIP wersja 8.2. Ułatwia on tworzenie i modyfikowanie jednostek stowarzyszonych, pomimo, że można do tego celu używać tych samych okien dialogowych, co w produkcie EIP w wersji 7.1 i wcześniejszych.

Aby utworzyć jednostkę stowarzyszoną za pomocą kreatora, wykonaj poniższe czynności:

1. **Zdefiniuj nazwę jednostki stowarzyszonej** i opisz ją. Można także określić, czy jest dla niej dopuszczalne użycie w wyszukiwaniu tekstowym.
2. **Zdefiniuj nazwy atrybutów stowarzyszonych** i zmodyfikuj je.
3. **Odwzoruj atrybuty stowarzyszone** na atrybuty rodzime. Dostępne są narzędzia pozwalające na pobranie spisu zasobów serwera, wybrania atrybutów rodzimych do modyfikowania oraz do późniejszego wprowadzania zmian w odwzorowaniach.
4. **Zdefiniuj właściwości** dla każdego atrybutu stowarzyszonego. Można dostosować właściwości atrybutów lub zaakceptować ustawienia domyślne.
5. **Potwierdź jednostkę stowarzyszoną**, przeglądając wcześniej ustawienia jednostki stowarzyszonej. Można także wrócić do wcześniejszych paneli w celu ich zmodyfikowania. Po dokonaniu zmian kliknij przycisk **Zakończ**.

Czynności te odpowiadają kolejnym krokom do wykonania w kreatorze. Więcej informacji na temat używania kreatora znajduje się w pomocy elektronicznej EIP.

Tworzenie stowarzyszonych indeksów tekstowych

Na serwerach Content Manager wersja 7.1 i wcześniejsza można zainstalować Text Search Engine, umożliwiając automatyczne indeksowanie, wyszukiwanie i pobieranie dokumentów przechowywanych w programie Content Manager. Dokumenty mogą być przeszukiwane przez podanie szukanych wyrazów lub całych fraz. Serwer wyszukiwania tekstowego obsługuje zestawy znaków jednobajtowych i dwubajtowych.

Jeśli serwery Content Manager są używane z mechanizmem wyszukiwania tekstowego w wersji 7.1 i wcześniejszej, użytkownik może utworzyć stowarzyszony indeks tekstowy i odwzorować go na indeks wyszukiwania tekstowego programu Content Manager na serwerach wyszukiwania tekstowego Content Manager.

Podczas tworzenia stowarzyszonego indeksu wyszukiwania tekstowego można włączyć opcję wyszukiwania złożonego, to znaczy przeszukiwania rodzimych indeksów tekstowych i atrybutów rodzimych. Włączenie dla wyszukiwania łączonego opcji stowarzyszonego indeksu tekstowego powoduje również odwzorowanie tego indeksu na jednostkę stowarzyszoną i odwzorowanie atrybutów rodzimych odwzorowanych przez tę jednostkę stowarzyszoną i jej atrybuty stowarzyszone na rodzime indeksy wyszukiwania tekstowego na serwerach wyszukiwania tekstowego.

Tworzenie szablonów wyszukiwania

Po utworzeniu jednostki stowarzyszonej można rozpocząć tworzenie szablonu wyszukiwania. Należy pamiętać, że szablon wyszukiwania wykorzystuje jednostkę stowarzyszoną jako odwzorowanie rzeczywistych danych. Przy tworzeniu szablonu wyszukiwania nadal należy zdefiniować, czego ma dotyczyć wyszukiwanie, co ma się stać z jego wynikami i kto ma mieć zezwolenie na korzystanie z szablonu. Dla każdego szablonu można używać tylko jednej jednostki stowarzyszonej, ale może ona zostać użyta w wielu szablonach. Można także wyszukiwać dowolną kombinację atrybutów jednostki stowarzyszonej. Aby utworzyć szablon wyszukiwania, można wykonać poniższe czynności przy użyciu kreatora szablonów wyszukiwania:

1. zdefiniować szablon wyszukiwania;
2. zdefiniować kryteria wyszukiwania;
3. zdefiniować ustawienia wyszukiwania;
4. przypisać uprawnienia dostępu.

Czynności te odpowiadają kolejnym krokom do wykonania w kreatorze. Szczegółowe informacje na temat procesu tworzenia szablonu umieszczone są w pomocy elektronicznej programu EIP.

Definiowanie szablonu wyszukiwania

Po uruchomieniu kreatora pojawi się okno umożliwiające rozpoczęcie definiowania szablonu wyszukiwania. Należy być przygotowanym na:

- Podanie nazwy i opisu dla szablonu wyszukiwania.
- Wybranie jednostki stowarzyszonej dla szablonu wyszukiwania. **Ograniczenie:** W szablonie wyszukiwania można użyć tylko jednej jednostki stowarzyszonej.
- Wybranie stowarzyszonego indeksu tekstowego, jeżeli jest to potrzebne.

Wskazówka: Pole wyboru stowarzyszonego indeksu tekstowego pojawi się tylko wtedy, jeśli z produktem Content Manager 7.1 lub wcześniejszym jest używany mechanizm Text Search Engine. Jeżeli do wyszukiwania tekstowego jest używany produkt DB2 TIE, jest to wyszukiwanie parametryczne i można je odpowiednio skonfigurować w szablonie wyszukiwania.

Definiowanie kryteriów wyszukiwania

Po zdefiniowaniu szablonu wyszukiwania kreator poprosi o:

1. Wybierz typ wyszukiwania: atrybutu lub dokumentu. Wyszukiwanie dokumentu jest dostępne tylko wtedy, gdy w poprzednim kroku został wybrany stowarzyszony indeks tekstowy.
2. podanie nazwy kryterium wyszukiwania;
3. wybranie atrybutu stowarzyszonego;
4. wybranie dostępnych operatorów;
5. podanie domyślnego poszukiwanego łańcucha (tylko dla wyszukiwania dokumentów).

Kreator ma rozwijane menu zawierające wszystkie atrybuty stowarzyszone powiązane z wybraną jednostką stowarzyszoną. Atrybuty te będą kryteriami wyszukiwania dla szablonu wyszukiwania. Kreator udostępnia także listę dostępnych operatorów.

Wskazówka: w szablonie można zdefiniować kilka kryteriów wyszukiwania albo usunąć z szablonu istniejące kryteria.

Określanie ustawień wyszukiwania

To okno pozwala na zdefiniowanie domyślnych ustawień wyszukiwania, ustawień dla kryteriów i ustawień dla wyświetlanych wartości. Każde z tych ustawień ma wartość domyślną, którą można zmieniać. W tym celu należy kliknąć odpowiedni przycisk dla każdego z nich.

Okno z domyślnymi ustawieniami pozwala na:

- Określenie, co ma się stać, jeśli serwer jest niedostępny w chwili gdy użytkownik aplikacji klienckiej chce użyć szablonu wyszukiwania.
- Definiowanie znaku zastępczego dla wyszukiwania parametrycznego.
- Określenie nazwy folderu do zapisania wyników wyszukiwania.

- Określenie, czy wyszukiwane ma być oparte na wszystkich (AND) czy dowolnym z (OR) kryteriów.

Okno ustawień kryteriów pozwala na kontrolowanie kolejności wyświetlania kryteriów wyszukiwania, kolejności kolumn wyników oraz nagłówków i szerokości tych kolumn.

Okno wyświetlanych wartości pozwala na zdefiniowanie sposobu wyświetlania wyników wyszukiwania. Na przykład, jeśli wartością parametru **Dzień tygodnia** na jednym serwerze jest **poniedziałek**, a na innym **pon.on.**, można określić, że dla obu serwerów ma być wyświetlany wynik wyszukiwania **poniedziałek**.

Przypisanie uprawnień

Po zdefiniowaniu, gdzie ma być prowadzone wyszukiwanie (przy użyciu jednostek stowarzyszonych), czego szukać (kryteria wyszukiwania), jak mają być wyświetlane wyniki (ustawienia), należy także zdefiniować, kto ma mieć dostęp do szablonów wyszukiwania.

Kreator udostępnia narzędzia pozwalające na przypisanie dostępu do szablonu dla wybranych użytkowników lub grup.

Przypisanie użytkownikowi praw dostępu do szablonu wyszukiwania nie jest równoznaczne z zezwoleniem mu na dostęp do każdego z serwerów danych przypisanych do szablonu.

Użytkownicy muszą spełniać wymagania dotyczące zabezpieczeń ustawione dla każdego serwera danych. Przed zezwoleniem użytkownikowi na dostęp do szablonu wyszukiwania należy użyć odpowiednich mechanizmów zarządzania użytkownikami i list kontroli dostępu aby upewnić się, że użytkownik ma odpowiednie uprawnienia.

Jeżeli kreator jest używany do wyszukiwania użytkowników lub grup, EIP zwraca tylko tych użytkowników, którzy mają odpowiednie prawa dostępu do żądanych serwerów danych.

Rozdział 4. Zarządzanie dostępem użytkowników

Użytkownicy nie mogą uzyskać dostępu do systemu EIP nie mając identyfikatora, hasła i zbioru uprawnień. Przed utworzeniem identyfikatorów i przypisaniu im uprawnień, należy zdefiniować, kto powinien mieć dostęp do systemu i jakie powinny przysługiwać tej osobie uprawnienia. Na przykład użytkownicy, którzy nie zdają sobie sprawy z konsekwencji usuwania obiektów, nie powinni mieć możliwości ich usuwania. Z drugiej strony nie powinno blokować się użytkownikom możliwości wykonywania ich pracy, nie nadając im poprawnych uprawnień. Tak więc przed nadaniem użytkownikom uprawnień, należy określić typy zadań wymaganych przez każdą pracę.

Gdy użytkownicy tworzą obiekt w systemie EIP, muszą zdefiniować prawa dostępu do tego obiektu dla innych użytkowników. Trzeba określić, kto może mieć dostęp do tego obiektu i jakie operacje może na nim wykonywać. Taka definicja praw dostępu nazywana jest w systemie EIP listą kontroli dostępu (ACL).

Tworzenie identyfikatorów i haseł użytkowników

Aby identyfikator użytkownika zdefiniowany w kliencie administracyjnym systemu był używany również przy uwierzytelnianiu DB2, identyfikator ten musi być zgodny z regułami nazewnictwa DB2. Reguły nazewnictwa DB2 dotyczą identyfikatorów użytkownika, które mają być używane dla superadministratorów lub dla użytkowników używanych do nawiązania połączenia. Nie można w nich używać następujących słów:

- USERS
- ADMINS
- GUESTS
- PUBLIC
- LOCAL
- Wszystkie zarezerwowane słowa SQL są wymienione w podręczniku SQL.

Identyfikator użytkownika nie może rozpoczynać się od następujących łańcuchów znaków:

- SQL
- SYS
- IBM

Możliwe jest używanie następujących znaków:

- Litery od A do Z **Ograniczenie:** Niektóre systemy operacyjne dopuszczają na stosowanie w identyfikatorach i hasłach wielkich i małych liter. Sprawdź w dokumentacji systemu operacyjnego, czy pozwala on na stosowanie wielkich i małych liter.
- Cyfry od 0 do 9.
- #
- \$

Ograniczenie: Długość identyfikatorów użytkownika nie może przekraczać 30 znaków.

Uprawnienia administracyjne bazy danych DB2

Po zalogowaniu się do klienta administracyjnego dostępne są dwa poziomy uwierzytelniania: jeden na poziomie bazy danych i drugi na poziomie produktu. Po włączeniu domen administracyjnych dostępne są też dwa rodzaje administratorów: superadministratorzy i podadministratorzy. W zasadzie tylko superadministratorzy mają dostęp do klienta administracyjnego systemu.

Superadministratorzy muszą mieć następujące uprawnienia DB2: db2admin, czyli muszą mieć pełne uprawnienia administracyjne do bazy danych DB2. Identyfikator użytkownika musi mieć zdefiniowane w systemie operacyjnym uprawnienia db2admin. Hasło użytkownika systemu operacyjnego jest używane do nawiązania połączenia z bazą danych DB2 i do zalogowania się do serwera bibliotecznego. Nie jest używane hasło zdefiniowane dla serwera bibliotecznego. Uprawnienia systemu Content Manager: Ten identyfikator użytkownika jest zdefiniowany w serwerze bibliotecznym z pełnymi uprawnieniami administracyjnymi systemu Content Manager ("AllPrivs"), aby umożliwić mu wykonywanie wszystkich czynności administracyjnych.

Podadministratorzy nie wymagają uprawnień DB2. Podadministratorzy zajmują się tylko wybranymi sekcjami serwera bibliotecznego i dlatego mogą się logować do klienta administracyjnego systemu używając jednego z dwóch wymienionych poniżej sposobów:

- Jeśli identyfikator użytkownika jest identyfikatorem użytkownika w systemie operacyjnym, do nawiązania połączenia z bazą danych DB2 i do zalogowania się do serwera bibliotecznego zostanie użyte hasło systemu operacyjnego.
- Jeśli identyfikator użytkownika nie jest identyfikatorem użytkownika w systemie operacyjnym, do nawiązania połączenia z bazą danych DB2 i do zalogowania się do serwera bibliotecznego zostanie użyty identyfikator i hasło podane w postaci zaszyfrowanej w pliku cmbfedenv.ini (dla systemu Enterprise Information Portal) lub cmbicmenv.ini (dla systemu Content Manager).

Więcej informacji na temat logowania się do serwera bibliotecznego znajduje się w następnej sekcji.

Podadministratorzy również potrzebują uprawnień w systemie EIP. Dla wszystkich czynności administracyjnych wykonywanych w poddomenach niezbędne są im uprawnienia administratora domeny.

Łączenie się z bazą danych DB2 przy użyciu plików INI

Każda pozycja w pliku INI zawiera nazwę serwera bibliotecznego i parę zaszyfrowany identyfikator i hasło używane przy łączeniu się z bazą danych DB2. Zaszyfrowane identyfikatory użytkowników i hasła są definiowane podczas instalowania produktu. Identyfikatory te muszą być inne niż identyfikatory administratorów. System Enterprise Information Portal używa pliku cmbfedenv.ini do łączenia się z bazą danych DB2, a system Content Manager używa pliku cmbicmenv.ini. Domyślny identyfikator użytkownika używany przy łączeniu się z bazą DB2 to ICMCONCT. Podczas instalacji hasła dla serwera bibliotecznego i menedżera zasobów są umieszczane w trzech miejscach. Plik cmbicmenv.ini zawiera identyfikator użytkownika i hasło używane do uzyskania dostępu do serwera bibliotecznego. System operacyjny definiuje dostęp do bazy danych, która przechowuje serwer biblieczny i menedżera zasobów. Identyfikator użytkownika i hasło dla menedżera zasobów znajdują się w pliku ICMRM.properties.

Jeśli jest używany plik INI, czyli identyfikator użytkownika nie występuje w systemie operacyjnym, podane w tym pliku identyfikatory użytkownika muszą istnieć w serwerze bibliotecznym.

Identyfikator używany do połączenia z bazą danych musi być zdefiniowany w serwerze bibliotecznym oraz w systemie operacyjnym. Nie są dla niego wymagane uprawnienia UserDB2Connect. Aby zmienić identyfikator używany do połączenia oraz hasło podane w pliku INI, w kliencie administracyjnym wybierz kolejno opcje **Narzędzia --> Zmień hasło/ID bazy danych**.

Zmiana serwera bibliotecznego i hasła administratora systemu w menedżerze zasobów

Jeśli chcesz zmienić hasło w menedżerze zasobów, musisz zmienić hasło używane przez serwer biblioteczny do logowania się w menedżerze zasobów oraz hasło administratora dla menedżera zasobów. **Ważne:** Przy zmianie hasła do logowania się w serwerze bibliotecznym oraz hasła administratora systemu w menedżerze zasobów, wykonaj kolejno przedstawione poniżej czynności:

1. Zaloguj się do klienta administracyjnego systemu.
2. Rozwiń drzewo Menedżera zasobów.
3. Kliknij dwukrotnie menedżera zasobów, którego chcesz zmodyfikować i rozwiń jego drzewo.
4. Kliknij opcję Definicje serwera i wybierz opcję Właściwości. Zostanie wyświetlone okno panelu serwera.
5. Zmień hasło w polu Hasło.
6. Kliknij przycisk OK.
7. Kliknij prawym przyciskiem myszy menedżera zasobów i kliknij opcję Właściwości. Zostanie wyświetlone okno właściwości menedżera zasobów.
8. Zmień hasło w polu Hasło i kliknij przycisk OK.

Zmiana haseł dostępu do bazy danych

Jeśli chcesz zmienić hasła dostępu do bazy danych, musisz zmienić hasło systemu operacyjnego używane do nawiązywania połączeń z bazą danych oraz plik ICMRM.properties, aby menedżer zasobów mógł zidentyfikować nowe hasło.

Aby zmienić hasło systemu operacyjnego dla połączenia z bazą danych, wykonaj następujące czynności:

1. W zależności od rodzaju używanego systemu operacyjnego, otwórz aplikację służącą do zarządzania użytkownikami i hasłami.
2. Kliknij użytkownika ICMRM.
3. Wybierz opcję zmiany hasła.
4. Wpisz nowe hasło.

Aby zmienić plik ICMRM.properties, wykonaj następujące czynności:

1. Otwórz plik ICMRM.properties. Jego domyślne położenie to X:\WebSphere\AppServer\installedApps\icrmr.ear\icrmr.war\WEB-INF\classes\com\ibm\mm\icrmr\ICMRM.properties, gdzie X jest dyskiem, na którym został zainstalowany system Content Manager.
2. Zmień hasło bazy danych DB2 tak, aby odpowiadało hasłu systemu operacyjnego.
3. Zapisz plik ICMRM.properties.

Po zmianie hasła bazy danych należy ją zrestartować albo pozwolić na wystąpienie dwóch lub trzech błędów, aż baza zresetuje się samoczynnie.

Szczegółowe informacje na temat zmiany haseł i innych ustawień menedżera zasobów przy użyciu klienta administracyjnego systemu można znaleźć w pomocy elektronicznej dla menedżera zasobów.

Importowanie użytkowników z katalogu LDAP

Protokół LDAP obsługuje zarządzanie identyfikatorami użytkowników i ich hasłami na poziomie całego przedsiębiorstwa, a nie pojedynczego systemu. System EIP może współpracować z trzema rodzajami serwerów LDAP: IBM Directory (poprzednio noszący nazwę IBM SecureWay Directory), Windows 2000 Active Directory oraz Lotus Domino Directory Notes Address Book (NAB). Hasło użytkownika jest przechowywane na serwerze LDAP. Gdy użytkownik loguje się do systemu lub Enterprise Information Portal, jego identyfikator, hasło oraz uprawnienia są sprawdzane w profilu użytkownika znajdującym się w bazie danych EIP. Opcja LDAP może zostać włączona podczas instalacji systemu EIP. Jeżeli protokół LDAP nie był włączony podczas instalacji, można go włączyć w dowolnym momencie.

Aby włączyć opcję LDAP, wybierz kolejno opcje **Start** → **Programy** → **EIP for Multplatforms** → **Program planujący import użytkowników LDAP**, a następnie uruchom klienta administracyjnego systemu. Wyświetli okno konfiguracyjne LDAP (Narzędzia --> Konfiguracja LDAP). Zaznacz pole wyboru Włącz importowanie użytkowników i uwierzytelnianie LDAP i podaj na stronie Serwer informacje o serwerze LDAP.

Po włączeniu obsługi LDAP będzie można importować użytkowników po kliknięciu przycisku LDAP w oknie Nowy użytkownik. Pozwoli to na selektywne importowanie użytkowników z serwera LDAP do systemu EIP. Można również importować całe grupy użytkowników używając programu planującego import użytkowników LDAP. Podczas logowania się serwer biblioteczny automatycznie połączy się z serwerem LDAP w celu uwierzytelnienia użytkownika. Jeśli serwer LDAP z jakiegóż przyczyny nie będzie w stanie zweryfikować hasła użytkownika, uwierzytelnienie zostanie zakończone niepowodzeniem.

Możesz zmodyfikować konfigurację serwera LDAP otwierając główne okno klienta administracyjnego i klikając kolejno opcje **Narzędzia** -> **Konfiguracja LDAP**. Możesz również zmienić bieżący serwer LDAP uruchamiając program narzędziowy do importowania użytkowników LDAP. W tym celu wybierz opcję Start (w systemie EIP). Informacje dotyczące planowania serwera LDAP zawiera podręcznik *Planning and Installing Your Content Management System*. Informacje na temat konfigurowania informacji serwera LDAP w oknach administracyjnych systemu znajdują się w pomocy elektronicznej dla systemu.

Informacje dotyczące planowania serwera LDAP zawiera podręcznik *Planning and Installing Your Content Management System*. Więcej informacji na temat implementowania LDAP znajduje się w pomocy elektronicznej programu klient administracyjny systemu.

Wprowadzenie do uprawnień

Klient administracyjny systemu udostępnia grupy uprawnień, zbiory uprawnień oraz pojedyncze uprawnienia. Jeśli administrujesz złożonym systemem Content Manager/EIP, uprawnienia są wspólne dla obu części klienta. Uprawnienia wbudowane w klienta mogą znacznie ułatwić tworzenie:

Grupy przywilejów

Grupa przywilejów jest kolekcją zadań użytkowników, która ma pomóc administratorom w utworzeniu nowych zbiorów uprawnień lub zastosowaniu uprawnień wyświetlanych w oknie Uprawnienia.

Zbiory uprawnień

Zbiory uprawnień są kolekcjami ról użytkowników.

Uprawnienie

Uprawnienia reprezentują czynności wykonywane przez użytkownika. Na przykład:

Przykład 1 - uprawnienia: Chcesz przydzielić uprawnienia ClientScan i ClientImport grupie użytkowników, którzy zazwyczaj wykorzystują klienta tylko do skanowania i importowania dokumentów do systemu Content Manager. Jeśli czynności te są zazwyczaj wykonywane przez wielu użytkowników, można utworzyć jeden identyfikator użytkownika, np. uzytkownik1. Następnie należy mu przypisać uprawnienia ClientScan i ClientImport. Następnie należy przypisać tego użytkownika do grupy grupa1. Jeśli dowolny użytkownik końcowy zaloguje się do swojego klienta jako uzytkownik1, uzyska dostęp tylko do skanowania i importowania dokumentów do systemu Content Manager.

Przykład 2 2 - grupy uprawnień: Istnieje grupa doświadczonych użytkowników, którzy wymagają dostępu do wszystkich typowych zadań wykonywanych przez klienta. Można dla nich utworzyć kolejny identyfikator użytkownika, np. uzytkownik2. Następnie można go przypisać do grupy np. o nazwie grupa2. Następnie do konta uzytkownik2 można przypisać grupę uprawnień o nazwie ClientTaskAll. Jeśli dowolny użytkownik końcowy zaloguje się do swojego klienta jako uzytkownik2, uzyska dostęp do wszystkich czynności dostępnych w grupie uprawnień o nazwie ClientTaskAll.

Przykład 3 - zbiory uprawnień: Pewna grupa użytkowników wymaga dostępu tylko do odczytu. Można dla nich utworzyć kolejny identyfikator użytkownika, np. uzytkownik3. Następnie można go przypisać do grupy np. o nazwie grupa3. Następnie do konta uzytkownik3 można przypisać grupę uprawnień o nazwie ClientUserReadOnly. Jeśli dowolny użytkownik końcowy zaloguje się do swojego klienta jako uzytkownik3, uzyska dostęp do wszystkich czynności dostępnych w grupie uprawnień o nazwie ClientUserReadOnly.

Tworzenie zbiorów uprawnień

Planując konfigurację systemu EIP należy określić, kto będzie mieć do niego dostęp i jakie będą nadawane prawa dostępu. Dostęp w systemie EIP jest definiowany za pomocą uprawnień.

Dzięki uprawnieniom można zezwolić na dostęp do określonego obiektu w ściśle określony sposób. Uprawnienia obejmują prawa do tworzenia, usuwania i wybierania obiektów przechowywanych w systemie. Zbiorem uprawnień nazywa się grupę uprawnień nadanych użytkownikowi.

Pierwszym zadaniem związanym z zarządzaniem prawami dostępu jest utworzenie zbiorów uprawnień dla użytkowników. *Zbiór uprawnień* określa zadania lub czynności, które dany użytkownik może wykonywać. Zbiory uprawnień łączą uprawnienia i są dostosowane do określonych typów użytkowników. Można na przykład określić jedną grupę administratorów zarządzających serwerami serwerem przepływu dokumentów, a drugą grupę zarządzających domenami. Gdy administrator zaloguje się, program EIP sprawdza jego zbiór uprawnień.

Program klient administracyjny systemu wyposażony jest w uprawnienia, które można łączyć w celu utworzenia zbioru uprawnień. Następnie zbiory uprawnień można przypisywać poszczególnym użytkownikom. Nie można przypisać zbioru uprawnień do grupy użytkowników.

Tworzenie grup uprawnień

Grupy uprawnień są podobne do grup użytkowników. Grupę uprawnień tworzy się w celu łatwiejszego znajdowania uprawnień mających tworzyć zbiór uprawnień. Na przykład, jeżeli

zdefiniowane są dwa uprawnienia, które przypisywane są prawie każdemu użytkownikowi w systemie, zamiast wyszukiwać ich za każdym razem, gdy tworzony jest zbiór uprawnień, wystarczy utworzyć grupę o nazwie BasicPrivs.

Przypisywanie użytkownikowi zbioru uprawnień

Program klient administracyjny systemu wyposażony jest w uprawnienia, które można łączyć w celu utworzenia zbioru uprawnień. Następnie zbiory uprawnień można przypisywać poszczególnym użytkownikom. Nie można przypisać zbioru uprawnień do grupy użytkowników.

Użytkownik może utworzyć nową nazwę uprawnienia, ale nie może utworzyć samego uprawnienia. Tylko programista systemowy może utworzyć uprawnienia, które nie są zdefiniowane w programie klient administracyjny systemu.

Można wykorzystać zbiory uprawnień, dostarczane wraz z programem EIP, albo utworzyć własne.

Przypisywanie użytkownikowi zbioru uprawnień nadawania

Aby uniemożliwić użytkownikom tworzenie nowych użytkowników, których uprawnienia są większe od tych, które mają w chwili tworzenia, program EIP implementuje zbiór uprawnień nadawania. Przypisanie użytkownikowi zbioru uprawnień nadawania oznacza, że użytkownik może tworzyć nowych użytkowników, ale nie może nadawać im uprawnień wyższych niż jego własne. Na przykład można nadać użytkownikowi uprawnienia administracyjne do zarządzania domeną. W niektórych sytuacjach może być przydatne uniemożliwienie tworzenia nowych użytkowników. Dlatego podczas tworzenia takiego użytkownika w polu Zbiór uprawnień nadawania należy wybrać "Noprivs". W rezultacie taki użytkownik będzie mógł zarządzać domeną, ale nie będzie mógł tworzyć w niej użytkowników.

Przypisywanie użytkowników do menedżera zasobów

Aby umożliwić użytkownikom dostęp do określonego menedżera zasobów, należy przypisać menedżer zasobów do domeny, do której użytkownicy mają dostęp. Więcej informacji na temat przypisywania menedżera zasobów do domeny znajduje się w sekcji "Przypisywanie menedżera zasobów do domeny" na stronie 39.

Przypisywanie użytkowników do kolekcji

Aby umożliwić użytkownikom dostęp do kolekcji, należy przypisać kolekcję na menedżerem zasobów do domeny, do której użytkownicy mają dostęp. Więcej informacji na temat przypisywania kolekcji do domen zawiera sekcja "Przypisywanie kolekcji do domeny" na stronie 39.

Tworzenie grup użytkowników

Często użytkownicy na takich samych stanowiskach wykonują takie same albo podobne zadania, tak więc powinni mieć taki sam dostęp do systemu. Użytkowników takich można połączyć w grupy. Nie można zagnieżdżać grup wewnątrz innych grup.

Grupowanie użytkowników jest opcją dostarczaną tylko dla wygody połączenia użytkowników o podobnych zadaniach. Nie można przypisywać grupom zbiorów uprawnień. Każdy użytkownik w grupie ma własny zbiór uprawnień. Grupowanie użytkowników ułatwia definiowanie list kontroli dostępu.

Jeżeli włączona jest obsługa domen, przed przypisaniem użytkownika do grupy należy sprawdzić, czy grupa ta jest umieszczona w określonej domenie, czy też w domenie

publicznej (więcej informacji na temat domen znajduje się w sekcji “Administrowanie domenami” na stronie 38). Należy upewnić się, czy grupa jest w domenie, w której ma być wybrany użytkownik. Jeżeli ma być utworzony użytkownik dla określonej domeny, to należy kliknąć **Nowy użytkownik** w oknie Grupa użytkowników. Następnie można dodać utworzonego użytkownika do grupy i upewnić się, że jest on w tej samej domenie.

Tworzenie list kontroli dostępu

Użytkownicy powinni posiadać uprawnienia wystarczające do wykonania ich obowiązków. Określanie dostępu do obiektów potraktowanych indywidualnie jest oddzielną grupą zagadnień.

Lista kontroli dostępu (ACL) zawiera jeden lub wiele identyfikatorów użytkowników lub grup użytkowników i przyznane im uprawnienia. Listy dostępu służą do sterowania dostępem użytkowników do obiektów w systemie EIP. Obiekty, które mogą mieć przypisaną listę kontroli dostępu, to: obiekty danych zapisane przez użytkowników, typy elementów i podzbiory typów elementów, listy prac i procesy.

Zbiory uprawnień definiują dla użytkownika indywidualne maksimum możliwości używania systemu, lista kontroli dostępu ogranicza indywidualnym użytkownikom dostęp do obiektu. Lista ACL, która ma uprawnienia nie zdefiniowane przez zbiór uprawnień użytkownika, nie zezwala na dostęp z tym uprawnieniem. Tylko użytkownicy, którzy mają to uprawnienie, mogą używać tego uprawnienia dla obiektu. Lista ACL ogranicza dostęp użytkowników, nie nadaje dodatkowych uprawnień dostępu. Lista kontroli dostępu dodaje dodatkowy poziom ochrony w systemie.

Przypisywanie zbiorów uprawnień do listy kontroli dostępu

Każdy użytkownik, dodany do listy kontroli dostępu, musi mieć przypisany zbiór uprawnień. Dzięki temu możliwe jest określenie, który użytkownik ma dostęp do określonego obiektu i jakie ma uprawnienia dostępu.

Użytkownicy nie mogą uzyskać dostępu do obiektu, który nie jest umieszczony na liście ACL. Aby dodać użytkownika lub grupę użytkowników do listy ACL, należy wybrać zbiór uprawnień dla listy ACL i kliknąć przycisk **Dodaj**. Dla każdej zdefiniowanej listy ACL można uzyskać listę użytkowników i grup. Wyświetloną tabelę można modyfikować, dodając i usuwając użytkowników i grupy. Więcej informacji na temat tworzenia i modyfikowania list ACL znajduje się w pomocy elektronicznej programu klient administracyjny systemu.

Tworzenie domen

Domena jest sekcją administracyjnej bazy danych, którą zarządza jeden lub więcej administratorów. Domeny składają się z identyfikatorów użytkowników, grupy użytkowników, zbiorów uprawnień, list kontroli dostępu, menedżerów zasobów i kolekcji SMS. Domeny nie są widoczne dla użytkowników, mają znaczenie wyłącznie dla administratorów, którzy będą nimi zarządzać. Użytkownicy nie wiedzą, że mają dostęp ograniczony tylko do części administracyjnej bazy danych, a zatem znają tylko elementy umieszczone w ich domenie.

Domeny administracyjne służą do ograniczenia administratorom i użytkownikom dostępu do poszczególnych sekcji administracyjnej bazy danych. Administratorzy mający pełne uprawnienia dostępu do administracyjnej bazy danych, mogą przekazać część uprawnień innym administratorom. Administrator z pełnymi uprawnieniami, superadministrator, ma dostęp do wszystkich sekcji administracyjnej bazy danych, natomiast zwykły administrator z ograniczonymi uprawnieniami ma dostęp tylko do pojedynczej sekcji.

Używanie domen ogranicza dostęp podadministratorom do list kontroli dostępu (ACL). Tylko główny administrator może tworzyć listy ACL, podadministratorzy mogą jedynie dodawać i usuwać z nich użytkowników i grupy użytkowników. Podadministratorzy nie mogą tworzyć, aktualizować lub usuwać list ACL.

Podadministrator może używać części uprawnień superadministratora, ale tylko w obrębie swojej domeny. Tworząc domeny i przypisując administratorów do tych domen, superadministrator może przekazać im mniej istotne zadania i skoncentrować się na całym systemie, wydajnie nim zarządzać, podczas gdy podadministratorzy mogą zarządzać użytkownikami i wykonywać określone czynności w domenie.

Przed włączeniem obsługi domen należy rozważyć następujące czynniki:

- Raz włączonych domen nie można wyłączyć.
- Menedżerowie zasobów, kolekcje, identyfikatory użytkowników i grupy użytkowników mogą istnieć tylko w jednej domenie na raz.
- Zbiory uprawnień i listy kontroli dostępu mogą istnieć w wielu domenach jednocześnie.
- Z wyjątkiem domeny publicznej PUBLIC, domeny nie pokrywają się.
- Dowolny obiekt utworzony w domenie superadministracyjnej nie może zostać przeniesiony niezależnie od tego, czy został utworzony przez system czy przez użytkownika.

Aby włączyć obsługę domen, z menu Plik wybierz opcję **Narzędzia —> Domeny administracyjne**, a następnie **Włącz domeny administracyjne**. Ponownie uruchom klienta administracyjnego dla domen, aby zmiany stały się widoczne. Więcej informacji na temat konfigurowania domen w administracyjnej bazie danych znajduje się w pomocy elektronicznej programu administracyjnego systemu.

Administrowanie domenami

W zależności od zbioru uprawnień podadministrator może zarządzać albo całą administracyjną bazą danych, albo określoną domeną. Administrator mający pełny dostęp do administracyjnej bazy danych nazywany jest superadministratorem. Podadministrator ma pełny dostęp do obiektów tylko w określonej domenie.

Każdy administrator ma możliwość tworzenia, pobierania, aktualizacji i usuwania obiektów w domenie (również dotyczy to użytkowników i kolekcji). Podadministratorzy mogą przeglądać oraz wydobywać obiekty umieszczone w ich domenie lub wyświetlać i wydobywać obiekty w domenie publicznej (PUBLIC), czyli współużytkowanej.

Dostęp do domen

Podadministrator nie może zmieniać domeny obiektu. Może jednak uzyskiwać dostęp do zawartości własnej domeny oraz wyświetlać i wydobywać wszystkie obiekty w domenie publicznej PUBLIC.

Superadministrator ma dostęp do wszystkich domen w administracyjnej bazie danych. Może też tworzyć obiekty i przypisywać je do domeny. Niektóre obiekty, takie jak zbiory uprawnień i listy ACL muszą być tworzone dla podadministratorów.

Podadministratorzy mogą tylko tworzyć, wydobywać, aktualizować i usuwać (CRUD) obiekty we własnej domenie.

Przypisywanie użytkowników do domeny

Podczas tworzenia użytkownika dostępna jest opcja dodania go do domeny albo pozostawienia w domenie domyślnej. Przypisanie użytkownika do domeny można również zmienić później we właściwościach użytkownika.

Użytkownik może mieć w danej chwili dostęp tylko do jednej domeny. Nie można dodać użytkownika do domeny publicznej PUBLIC.

Tylko superadministrator ma możliwość tworzenia domen i przypisywania do nich użytkowników. W domenie może istnieć kilku podadministratorów, lecz mogą być oni tworzeni wyłącznie przez superadministrатора, przez nadanie im odpowiednich uprawnień. W polu **Nadawany zbiór uprawnień** w oknie Nowy użytkownik lub Właściwości użytkownika umieszczona jest informacja, jakie uprawnienia administracyjne ma podadministrator w domenie.

Przypisywanie grupy użytkowników do domeny

Przypisanie grupy użytkowników do domeny spowoduje zmianę przypisanej domeny dla każdego użytkownika w grupie. Użytkownik może mieć w danej chwili dostęp tylko do jednej domeny. Tak więc każdy użytkownik umieszczony w przypisywanej grupie jest przenoszony do nowej domeny.

W danej chwili grupa użytkowników nie może być tylko w jednej domenie. Grupa może być przypisana do domeny publicznej PUBLIC.

Przypisywanie zbioru uprawnień do domeny

Każdy identyfikator użytkownika, który jest dodawany do domeny, musi mieć przypisany zbiór uprawnień. Jeżeli użytkownikom nie zostaną przypisane uprawnienia, nie będą mogli oni wykonywać swoich obowiązków. Najlepszym miejscem do przechowywania zbiorów uprawnień w celu udostępniania ich użytkownikom jest domena publiczna (PUBLIC), czyli współużytkowana.

Przypisywanie menedżera zasobów do domeny

Można ograniczyć dostęp użytkowników do określonych serwerów obiektów przez przypisanie ich do wybranej domeny. Podczas definiowania nowego menedżera zasobów dla administracyjnej bazy danych dostępna jest opcja wyboru domeny.

Domyślną domeną dla wszystkich serwerów obiektów jest domena publiczna PUBLIC. Aby nie wszyscy mieli dostęp do menedżera zasobów, należy przypisać go do domeny. Jeżeli domena, do której może być przypisany menedżer zasobów, nie jest widoczna, wciąż można zdefiniować menedżera zasobów, a następnie utworzyć potrzebną domenę. Po zdefiniowaniu odpowiedniej domeny należy otworzyć menedżera obiektów i wybrać domenę.

Przypisywanie kolekcji do domeny

Można ograniczyć dostęp użytkowników do określonych kolekcji menedżera zasobów przez przypisanie ich do wybranej domeny. Jeżeli menedżer zasobów jest umieszczony w domenie publicznej PUBLIC, kolekcję można przypisać do dowolnej innej zdefiniowanej domeny. Jeżeli menedżer zasobów jest przypisany do określonej domeny, nie można przypisywać kolekcji do innych domen, nawet do domeny publicznej PUBLIC.

Aby użytkownik miał dostęp do kolekcji w menedżerem zasobów, musi on mieć do niego dostęp, dlatego nie można ograniczyć dostępu do menedżera zasobów bez nakładania tego samego ograniczenia na znajdujące się w nim kolekcje.

Przenoszenie użytkownika do innej domeny

Mogą istnieć powody do usunięcia użytkownika z jednej domeny i dodania go do innej. Aby pamiętać, w których grupach znajduje się użytkownik, można użyć pola **Opis** w oknie Definicja użytkownika. Może to ułatwić pracę.

Ważne: czynność ta zajmuje dużo czasu i nieprawidłowo wykonana może spowodować problemy z dostępem do systemu. Tylko superadministrator może zmienić domenę użytkownika.

Uważnie wykonaj następujące czynności:

1. Znajdź wszystkie grupy, do których należy użytkownik.
2. Albo przenieś wszystkie te grupy do domeny PUBLIC, albo usuń użytkownika ze wszystkich tych grup.
3. Przenieś każdy menedżer zasobów powiązany z tym użytkownikiem do domeny PUBLIC, a następnie wszystkie kolekcje każdego menedżera zasobów przenoszonego do domeny docelowej.
4. Utwórz, *nie przenoś* wszystkie zbiory uprawnień powiązane z użytkownikiem w domenie docelowej, chyba że już się tam znajdują.
5. Utwórz, *nie przenoś* wszystkie listy kontroli dostępu powiązane z użytkownikiem w domenie docelowej, chyba że już się tam znajdują.
6. Przenieś użytkownika do domeny docelowej, otwierając jego Właściwości i zmieniając domenę.
7. **Opcjonalnie:** można przenieść grupy i menedżerowie zasobów przeniesione w krokach 1, 2 i 3 z domeny PUBLIC do domeny docelowej, ale tylko pod warunkiem, że w domenie źródłowej nie ma już żadnych użytkowników powiązanych z przenoszonymi grupami i menedżerami zasobów. W przeciwnym razie grupy i menedżerowie zasobów muszą pozostać w domenie PUBLIC, aby umożliwić współużytkowanie dla użytkowników w różnych domenach.

Przypomnienie: w żadnym momencie użytkownik nie może znajdować się w domenie PUBLIC. Użytkowników nie można współużytkować.

Przenoszenie grupy użytkowników do innej domeny

Ważne: czynność ta nieprawidłowo wykonana może spowodować problemy z dostępem do systemu. Tylko superadministrator może zmienić domenę grupy użytkowników.

Aby przenieść grupę użytkowników do innej domeny:

- Jeśli grupa użytkowników jest pusta, usuń ją z bieżącej domeny, a następnie ponownie utwórz i przypisz do domeny docelowej.
- Jeśli grupa użytkowników nie jest pusta:
 1. Znajdź wszystkich użytkowników należących do tej grupy.
 2. Usuń grupę z bieżącej domeny, co spowoduje usunięcie wszystkich użytkowników.
 3. Ponownie utwórz grupę i przypisz ją do domeny docelowej.
 4. Dodaj wszystkich użytkowników do utworzonej teraz grupy.

Przenoszenie menedżera zasobów do innej domeny

Tylko superadministrator może zmienić domenę menedżera zasobów.

Aby przenieść menedżera zasobów do innej domeny:

- Jeśli menedżer zasobów nie zawiera kolekcji, przenieś go do domeny docelowej, otwierając jego właściwości i zmieniając domenę na docelową.

- Jeśli menedżer zasobów zawiera kolekcje:
 1. Przenieś menedżera zasobów do domeny PUBLIC.
 2. Przenieś kolekcje do domeny docelowej, otwierając Właściwości i wybierając domenę docelową.
 3. Przenieś menedżera zasobów do domeny docelowej, otwierając Właściwości i wybierając domenę docelową.

Przenoszenie kolekcji do innej domeny

Tylko superadministrator może zmienić domenę kolekcji.

Aby przenieść kolekcję do innej domeny:

1. Znajdź menedżera zasobów, do którego należy kolekcja.
2. Przenieś powiązanego menedżera zasobów do domeny PUBLIC.
3. Przenieś kolekcję do domeny docelowej, otwierając Właściwości i wybierając domenę docelową.
4. Przenieś menedżera zasobów do domeny docelowej, otwierając Właściwości i wybierając domenę docelową.

Przenoszenie zbioru uprawnień do innej domeny

Ponieważ zbiory uprawnień mogą znajdować się w wielu domenach, można dodawać je do domeny docelowej bez ich przenoszenia.

Przenoszenie listy kontroli dostępu do innej domeny

Ponieważ listy kontroli dostępu mogą znajdować się w wielu domenach, można dodawać je do domeny docelowej bez ich przenoszenia.

Rozdział 5. Zarządzanie eksploracją informacji

Niniejszy rozdział rozpoczyna się od wyjaśnienia, co to jest Eksploracja informacji i jak można ją wykorzystać w środowisku biznesowym. W dalszych sekcjach opisano pierwsze kroki z programem Eksploracja informacji, koncepcję działania programu Information Structuring Tool, a na zakończenie umieszczono kilka uwag dotyczących strojenia wydajności.

Co to jest Eksploracja informacji?

Enterprise Information Portal Eksploracja informacji jest kluczową techniką pozwalającą przedsiębiorstwom na zapewnienie użytkownikom łatwego i taniego dostępu do istotnych informacji dzięki zautomatyzowaniu wielu aspektów wyszukiwania i analizy informacji.

Podstawowym wyzwaniem pojawiającym się przy eksploracji informacji jest umożliwienie dostępu do danych przechowywanych w komputerach w postaci tekstu bez określonej struktury. Obecna technologia jeszcze nie pozwala na interpretację wiedzy zapisanej w postaci języka naturalnego. Istnieją natomiast narzędzia zdolne do pobierania wartościowych informacji z dowolnego pozbawionego struktury tekstu przy użyciu technik rozpoznawania wzorców i mechanizmów heurystycznych. Pobierane informacje mogą być różnego typu, od identyfikowania tak zwanych istotnych słów, takich jak nazwiska, instytucje lub miejsca, do tworzenia streszczeń dokumentów.

Eksploracja informacji nie polega jedynie na pobieraniu fragmentów informacji z pojedynczych dokumentów. Dobry moment na zastosowanie eksploracji informacji pojawia się wtedy, gdy mamy do czynienia z olbrzymią ilością dokumentów. Określenie "eksploracja informacji" (ang. "data mining") jest używane zarówno do procesu odkrywania wiedzy, czyli pobierania skodyfikowanych informacji z pojedynczych dokumentów i zapisywania ich jako metadanych, jak i do procesu analizy rozkładu tych cech w kolekcjach dokumentów w celu wykrycia interesujących zjawisk, wzorców lub trendów.

Usługi eksploracji informacji w produkcie Enterprise Information Portal

Usługa eksploracji informacji Enterprise Information Portal udostępnia infrastrukturę pozwalającą na tworzenie i zarządzanie informacjami związanymi z pojedynczymi dokumentami lub kolekcjami dokumentów. Informacje te są nazywane metadanymi. Przykłady informacji kategoryzujących dane i zapisanych jako metadane:

- tytuł,
- streszczenie,
- nazwy, terminy, wyrażenia,
- kategorie, do których należy dokument (np. oferta).

Cechą odróżniającą eksplorację informacji od tradycyjnej składnicy metadanych, która kojarzy dokumenty z metadanymi, jest możliwość automatycznego tworzenia metadanych nawet wtedy, gdy nie są one bezpośrednio dostępne. Algorytmy eksploracji i pobierania potrafią uzyskać istotne informacje z wielkich zbiorów dokumentów, bądź to kierując procesem eksploracji przy użyciu metadanych, bądź stosując dla metadanych modele statystyczne w celu znalezienia interesujących relacji, które mogą nie być widoczne przy rozpatrywaniu pojedynczych dokumentów należących do zbioru.

Szybkość procesu eksploracji informacji można znacząco podnieść dzięki przechowywaniu metadanych w dedykowanej składnicy, nazywanej *składnicą danych Eksploracji informacji*. Przyspiesza ona dostęp do metadanych, ponieważ aplikacja może pobierać dane z pojedynczego repozytorium i nie musi łączyć się ze zdalnymi serwerami danych. Ponieważ metadane mają kluczowe znaczenie dla nawigowania w zbiorze danych i eksploracji informacji oraz są one często używane do zawężania kryteriów wyszukiwania, można wykonywać wiele niezbędnych w tym celu czynności bez odwoływania się do serwerów danych.

Kolejną zaletą stosowania składnicy metadanych jest odseparowanie danych związanych z dokumentem od samego dokumentu. Dla wielu dokumentów, które są dostępne tylko do odczytu, tak jak zdalne dokumenty w sieci WWW, przechowywanie danych i metadanych w tym samym repozytorium w ogóle nie jest możliwe.

Eksploracja informacji produktu Enterprise Information Portal udostępnia wymienione poniżej mechanizmy służące do automatycznego tworzenia metadanych. Są to:

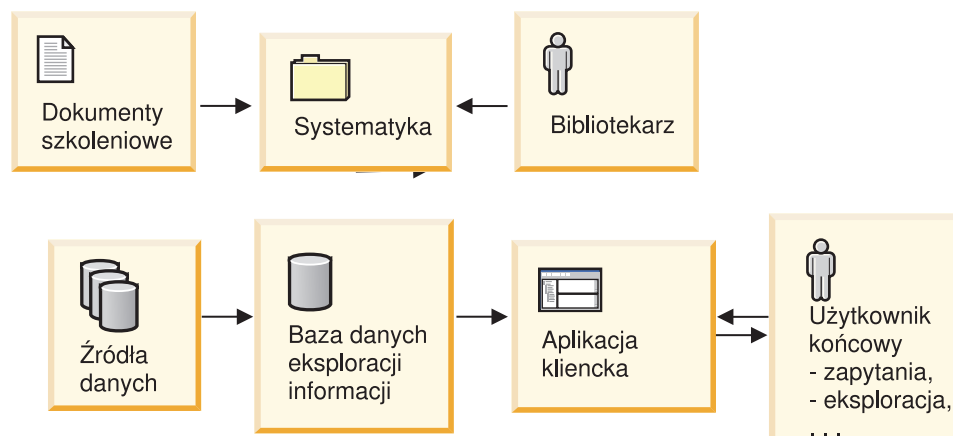
- **Klasyfikowanie** pozwala na przypisanie jednej lub kilku kategorii do dokumentu w oparciu o zdefiniowaną przez użytkownika systematykę. Składnik klasyfikujący zawiera aplikację udostępniającą graficzny interfejs użytkownika, pozwalający na tworzenie i zarządzanie systematykami, która nosi nazwę *Information Structuring Tool*.
- **Streszczanie** pozwala na pobieranie najistotniejszych zdań z dokumentu, co może pomóc w podjęciu decyzji, czy cały dokument ma być czytany, czy nie. Użytkownik może następnie określić długość żadanego podsumowania i w ten sposób ma bezpośredni wpływ na zależność między złożonością wyodrębnionych metadanych a ilością danych w dokumencie.
- **Identyfikacja języka** określa, w jakiej wersji językowej jest napisany dany dokument. Jest to użyteczna czynność wstępna, wykonywana przed innymi usługami.
- **Uzyskiwanie informacji** automatycznie rozpoznaje ważne terminy słownikowe w dokumentach tekstowych.
- **Łączenie w klastry** pozwala na utworzenie z jednego zbioru dokumentów wielu grup, zawierających podobne dokumenty. Klastry są tworzone automatycznie z kolekcji dokumentów.

Składniki usług eksploracji informacji

Aplikacja używająca eksploracji informacji zwykle wykonuje następujące zadania:

1. organizowanie danych w taki sposób, aby były przeglądane i nawigowane;
2. uzyskiwanie dostępu do różnych źródeł danych;
3. używanie zaawansowanych operacji wyszukiwania w celu odfiltrowania wymaganych danych niezbędnych do określania trendów lub prognoz.

Rys. 1 na stronie 45 ilustruje te zadania.



Rysunek 1. Zadania eksploracji informacji

Aby wykorzystać funkcjonalność eksploracji informacji, dokumenty muszą być tak organizowane, aby można je było przeglądać i nawigować wśród nich. Zadanie to jest zwykle wykonywane przez bibliotekarza lub inżyniera wiedzy. Bibliotekarz używa programu Information Structuring Tool (IST) w celu określenia systematyki i hierarchicznej charakterystyki tematycznej danych w dokumencie. IST jest aplikacją z graficznym interfejsem użytkownika służącą do tworzenia i zarządzania systematykami. Po przeszkoleniu kategorii i ustabilizowaniu się systematyki można uzyskiwać dostęp do różnych źródeł danych.

Dokumenty można importować z wielu źródeł danych EIP lub z sieci WWW za pomocą usługi Nawigator sieci WWW (dostępnej jedynie jako komponent JavaBeans), a następnie przypisywać je do kategorii.

Analiza tekstu i funkcje tworzenia metadanych są dostępne jako interfejs programistyczny (API) z poziomu niewizualnych komponentów JavaBeans i API usług Java.

Komponenty JavaBeans Eksploracja informacji to oprogramowanie służące do szybkiego tworzenia aplikacji zgodnych z JavaBeans. API usług Java zawiera pełną funkcjonalność usługi Eksploracja informacji dostępną podczas tworzenia własnych aplikacji. Dostępne są również przykłady stosowania komponentów JavaBeans wraz z przykładowymi stronami JSP.

Identyfikowanie zawartości dokumentu jest wymaganiem wstępnym dla wszystkich operacji wykonywanych przy użyciu eksploracji informacji, które przetwarzają zawartość dokumentu. Czynność ta obejmuje następujące operacje:

1. identyfikowanie strony kodowej dokumentu;
2. identyfikowanie bloków tekstu, które mają zostać przetworzone, tj. ignorowanie znaczników lub danych binarnych, takich jak obrazki;

Ponieważ dokumenty na serwerach danych mogą mieć arbitralnie wybraną strukturę, usługi Eksploracja informacji umożliwiają pisanie specyficznych modułów służących do identyfikowania i wyodrębniania istotnych bloków tekstu z dokumentów w różnych formatach. Usługa Eksploracja informacji zawiera domyślny moduł, który obsługuje szeroką gamę spotykanych formatów dokumentów. Więcej informacji na temat listy obsługiwanych formatów zawiera Rozdział 9, "Formaty dokumentów", na stronie 121, a informacje na temat używania przykładowego modułu domyślnego publikacja *Application Programming Guide*.

Tworzenie metadanych dla każdego wybranego dokumentu polega na przetwarzaniu treści dokumentu i zastosowaniu metod statystycznych lub heurystyki opartej na zasobach wiedzy, takich jak słowniki lub profile częstotliwości.

API eksploracji informacji obsługuje następujące operacje:

- streszczanie;
- klasyfikowanie;
- identyfikację języka;
- pobieranie informacji;
- klastrowanie.

Utworzone metadane dla wszystkich dokumentów są przechowywane w składnicy danych eksploracji informacji.

Po wypełnieniu składnicy danych dostępna staje się jeszcze jedna opcja wybierania dokumentów. Jest to opcja wybierania dokumentów w oparciu o informacje przechowywane w składnicy danych. Wyszukiwanie zaawansowane łączy w sobie zapytanie tekstowe z kategorią, ograniczając wyszukiwanie dokumentów do określonej kategorii.

Używanie eksploracji informacji w środowisku biznesowym

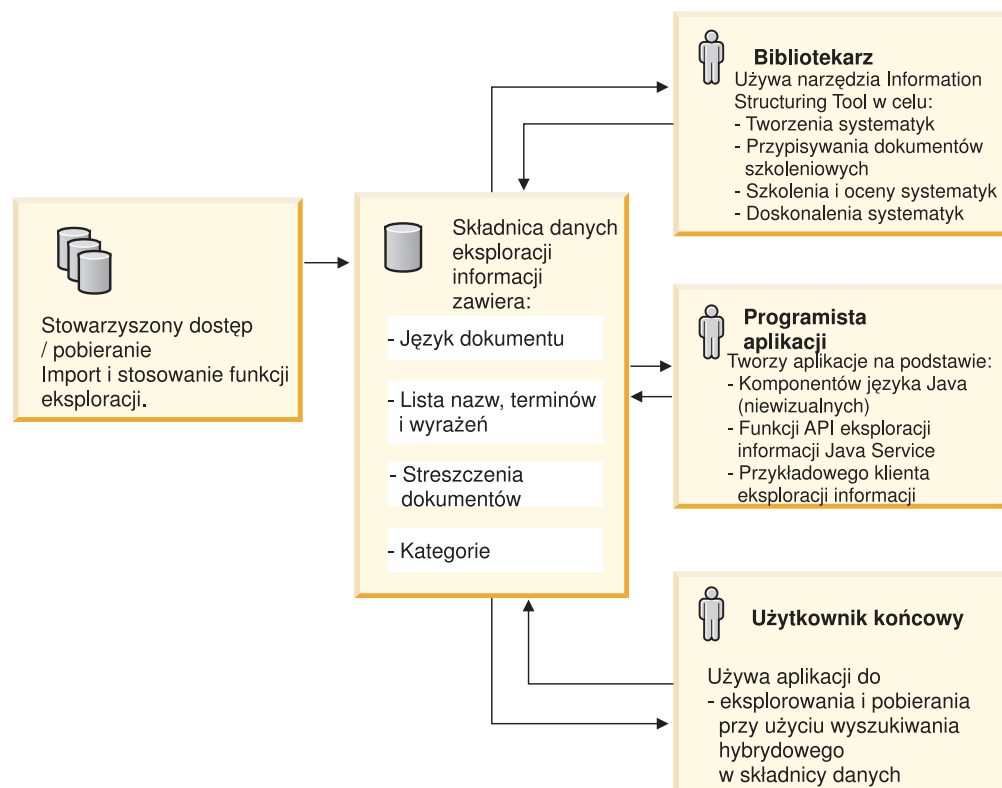
Infrastruktura organizacji, która implementuje dowolną technologię eksploracji danych, zwykle zawiera przynajmniej te role, które zostały wymienione poniżej:

- Administrator systemu informatycznego, który zajmuje się nie tylko eksploracją informacji.
- Programista aplikacji.
- Bibliotekarz lub inżynier wiedzy.
- Osoby pracujące z aplikacją obsługująca eksplorację danych (użytkownicy końcowi).

W zależności od natury aplikacji mogą również pojawić się dodatkowe, bardziej szczegółowe role:

- Projektant serwisów WWW.
- Architekt lub konsultant.

Rys. 2 na stronie 47 ilustruje te role i czynności.



Rysunek 2. Role i czynności w eksploracji informacji

Administrator systemu konfiguruje środowisko sprzętowe i programowe oraz zajmuje się utrzymaniem ich zasobów, np. przestrzenią dyskową i prawami dostępu. Instaluje także wymagane komponenty produktu Enterprise Information Portal i konfiguruje źródła danych oraz aplikację administracyjną tego produktu, umożliwiając dostęp do różnych repozytoriów. Administrator systemu administruje także składnicą danych na poziomie bazy danych.

Programista aplikacji tworzy aplikację, używając komponentów JavaBeans lub interfejsu API. Przykładowe programy JavaBeans zawiera publikacja *Application Programming Guide for Windows*, a opis funkcji API publikacja *Application Programming Guide*. Podczas tworzenia aplikacji pomocny może okazać się program *Web designer*.

Bibliotekarz lub *inżynier wiedzy* jest odpowiedzialny za konfigurowanie i utrzymywanie kolekcji dokumentów i zasobów, które będą używane do eksploracji i pobierania. Bibliotekarz używa aplikacji administracyjnej EIP do tworzenia przypisań metadanych i szablonów wyszukiwania oraz narzędzia Information Structuring Tool (patrz sekcja "Tworzenie systematyki" na stronie 63) do definiowania katalogów i systematyk. Zwykle do jego obowiązków należy także zapełnianie składnicy danych produktu Eksploracja informacji dokumentami pochodzącymi ze źródeł umieszczonych w sieci WWW lub innych źródeł danych, używając do tego celu aplikacji napisanych przez programistę.

Użytkownicy końcowi korzystają z aplikacji stworzonej przez programistę aplikacji do eksploracji informacji i pobierania ich przy wykorzystaniu zasobów, które są tworzone i utrzymywane przez bibliotekarza lub inżyniera wiedzy. W zależności od podziału pracy między użytkownikami końcowymi a bibliotekarzem, mogą oni także brać udział w wybieraniu dokumentów umieszczonych na serwerach danych i zapełnianiu składnicy metadanych produktu Eksploracja informacji.

Przykład zastosowania produktu Eksploracja informacji

Firma Elektryk produkuje urządzenia elektroniczne powszechnego użytku. W swojej ofercie posiada przynajmniej pięć różnych produktów, które mogą być skonfigurowane na bardzo różne sposoby.

Departament sprzedaży ma informacje o preferencjach użytkowników dotyczące określonych obszarów zastosowań. Te profile zastosowań opisują sposób korzystania z produktów i ich opcji konfiguracyjnych przez użytkowników. Każdemu z profili odpowiada określona strategia marketingowa oraz strategia dotycząca sprzedaży i zarządzania relacjami z klientem.

Departament usług posiada informacje dotyczące części, z których składają się produkty, sposobu ich łączenia, ich dostawców oraz niezawodności i serwisowania.

Administratorzy kontraktów przechowują informacje na temat sprzedawców i kooperantów. Mają także dostęp do dokumentów prawnych określających warunki realizacji poszczególnych umów.

W ostatnim czasie zanotowano spadek sprzedaży w określonych grupach klientów. Dotychczasowe sposoby korzystania z urządzeń elektronicznych uległy zmianie, zmieniły się również oczekiwania klientów, którzy chcą korzystać z nowości technologicznych.

Aby zareagować na te zmiany i wykorzystać je w celu osiągnięcia korzyści, firma Elektryk utworzyła grupę zadaniową, której celem jest opracowanie strategii pozwalającej firmie powrócić na rynek.

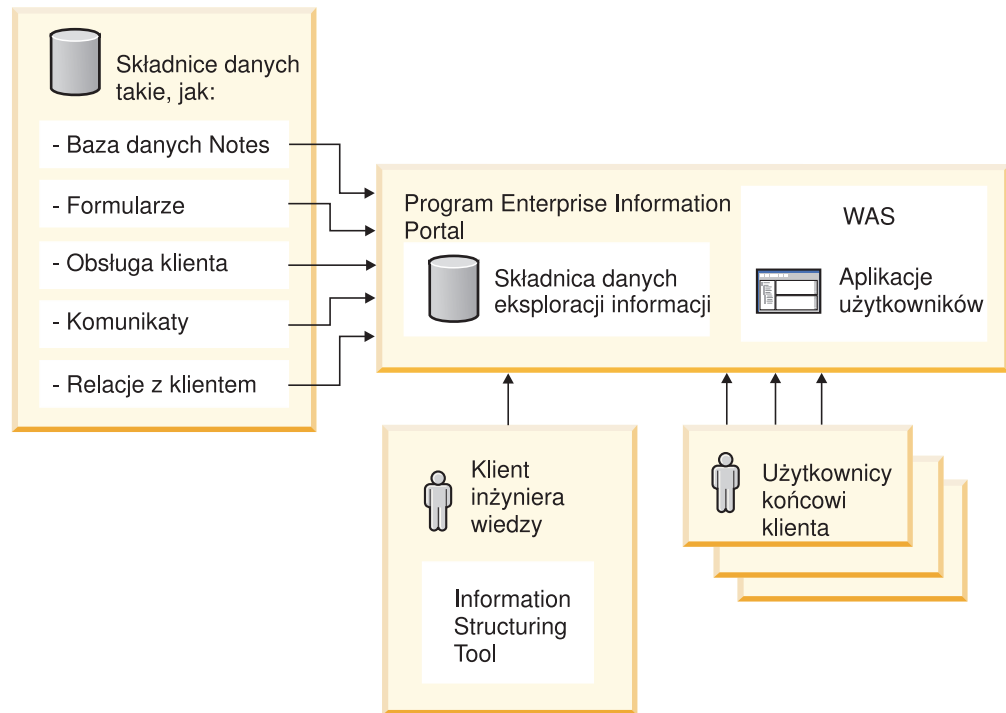
Pierwszym krokiem jest zaprojektowanie infrastruktury informatycznej, która umożliwi dostęp do istotnych informacji, pozwalając osobom tworzącym plany na podejmowanie szybkich decyzji w oparciu o pełne informacje.

Przykładami takich informacji są:

- Dane o produktach konkurencji, ich cechach, cenach i przyjmowaniu przez klientów.
- Wiedza o silnych i słabych stronach produktów firmy Elektryk widzianych z perspektywy klienta.
- Trendy rynkowe i perspektywy rozwoju obszarów, w których te urządzenia są zwykle stosowane.

Tego rodzaju informacje znajdują się w rozproszonych źródłach danych, na różnych platformach sprzętowych i programowych, a także są zorganizowane w różny sposób. Mogą to być informacje hierarchiczne, indeksowane lub nie, mogą także znajdować się w dokumentach różnego typu, np. w rekordach bazy danych lub w plikach HTML.

Rys. 3 na stronie 49 ilustruje używanie programu Eksploracja informacji.



Rysunek 3. Przykład eksploracji informacji

Grupa zadaniowa zdecydowała o konieczności utworzenia portalu udostępniającego wszystkie te rozproszone dane w taki sposób, aby można było uzyskać do nich dostęp przy użyciu przeglądarki WWW zainstalowanej na komputerze planisty. Wybrali produkt Enterprise Information Portal, ponieważ zawiera on wszystkie elementy składowe niezbędne do utworzenia takiego portalu, a nawet pozwala na utworzenie własnych łączników do wcześniejszych źródeł danych, które powstały w czasach przed powstaniem standardów dla aplikacji biznesowych.

Niezbędne było wykonanie następujących czynności:

1. Skonfigurowanie infrastruktury sprzętowej i programowej.
2. Zdefiniowanie metod dostępu do źródeł danych, skonfigurowanie niezbędnych połączeń i utworzenie przypisań do istotnych danych.
3. Zorganizowanie tych danych z taki sposób, aby planiści mogli łatwo przeglądać te dane i nawigować wśród nich.
4. Stworzenie aplikacji dla użytkownika końcowego.

W kroku 3 niezbędne jest zastosowanie produktu Eksploracja informacji. Po przygotowaniu i uruchomieniu infrastruktury zostały zidentyfikowane źródła danych, nawiązano odpowiednie połączenia i utworzono przypisania. Pozwoliło to na uzyskanie dostępu do żądanych danych z jednego miejsca. Podzbiory tych danych można definiować przy użyciu wyszukiwania stowarzyszonego. Następne pytanie brzmi: w jaki sposób można filtrować uzyskane dane, aby na ich podstawie uzyskać prognozy i trendy i jak te dane zorganizować, aby dostosować je do procesu planowania strategicznego.

Grupa zadaniowa zdefiniowała *inżyniera wiedzy*, który jest odpowiedzialny za zarządzanie, organizowanie i aktualizowanie informacji używanych przy strategicznym planowaniu. Aby wyodrębnić istotne informacje z dużego zbioru dokumentów znajdujących się w różnych źródłach danych, inżynier wiedzy rozmawia z osobami zajmującymi się poprzednio

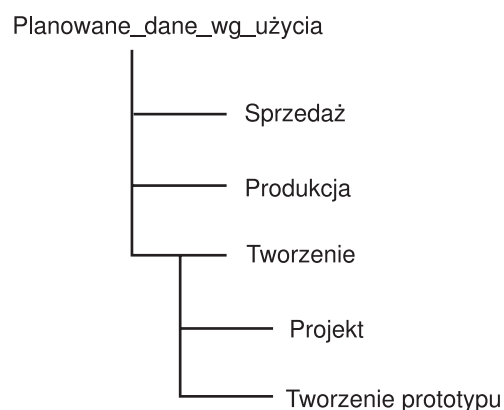
planowaniem strategicznym, aby dowiedzieć się, w jaki sposób procesy wzajemnie na siebie wpływają i co zostało uznane za dobrą praktykę. Następnie przeszukuje bazy danych zawierające informacje o relacjach z klientami.

Korzystając z funkcji przeszukiwania produktu Enterprise Information Portal, dokumenty z tych baz można łatwo przeglądać według nazwiska klienta, jego adresu lub według właściwości urządzenia. Informacja, którą trzeba odnaleźć w celu utworzenia profili zastosowania, jest ukryta między wierszami tekstu i jedynym sposobem dotarcia do niej jest analizowanie dokumentów w inteligentny sposób, stosując usługi eksploracji informacji produktu Enterprise Information Portal.

System ten dostarcza przydatnych informacji dzięki tematycznemu charakteryzowaniu zawartości dokumentu przez przypisywanie mu kategorii, np. *ten dokument dotyczy ogrodnictwa*. Usługa klasyfikowania produktu Eksploracja informacji przypisuje dokumentom kategorie analizując ich treść. Kategorie są powiązane w hierarchiczną strukturę zwaną systematyką. Zarówno dostępne bezpośrednio, jak i tworzone automatycznie metadane są umieszczone w repozytoriach, zarządzanych przez eksplorację informacji, nazywanych katalogami, które przyspieszają dostęp i pobieranie danych.

Przy użyciu narzędzia Information Structuring Tool inżynier wiedzy definiuje katalog, który ilustruje sposób korzystania z urządzeń przez klientów.

Rys. 4 na stronie 50 przedstawia katalog.



Rysunek 4. Przykład katalogu

Wykonując w bazach danych wspomagających sprzedaż i serwis wyszukiwanie stowarzyszone dla wszystkich klientów i sposobów korzystania przez nich z urządzeń, inżynier wiedzy może zidentyfikować zbiory dokumentów szkoleniowych, które są reprezentatywne dla każdej z kategorii w systematyce. Ta czynność może w efekcie prowadzić do przeorganizowania systematyki, ponieważ po bliższym zapoznaniu się z danymi mogą pojawić się nowe kategorie.

Gdy systematyka stanie się stabilna i do każdej kategorii będzie przypisana wystarczająca liczba dokumentów przykładowych, inżynier wiedzy może zacząć szkolenie systematyki za pomocą narzędzia Information Structuring Tool. Podczas szkolenia tworzony jest model klasyfikacji, który może być używany do przypisywania dokumentów do kategorii za pomocą usługi Klasyfikowania.

W tym czasie programiści z departamentu informatyki tworzą dla planistów aplikację wykorzystującą uszczuplonego klienta oraz komponenty JavaBeans lub interfejs API Java Service.

Aplikacja ta składa się z szeregu szablonów wyszukiwania dostosowanych do używania przy planowaniu strategicznym. Przy użyciu szablonów planiści mogą uzupełnić katalog dokumentami pobranymi z różnych źródeł. Podczas uzupełniania katalogu usługa Eksploracja informacji automatycznie przypisuje dokumentom kategorie. Jeżeli zostanie zidentyfikowany nowy profil zastosowań, inżynier wiedzy reorganizuje odpowiednio systematykę, używając jako materiału szkoleniowego nowych dokumentów zidentyfikowanych przez planistów. Katalog zostaje ponownie poddany szkoleniu i do planistów zostają przesłane nowe informacje.

Powyższy przykład jest ilustracją możliwości wykorzystania funkcjonalności produktu Eksploracja informacji w celu wyjścia naprzeciw potrzebom i oczekiwaniom klientów, co w efekcie pozwoli firmom, takim jak Elektryk, na utrzymanie konkurencyjności i pozostanie na rynku.

Obsługiwane języki i formaty

Usługi eksploracji informacji produktu Enterprise Information Portal Eksploracja informacji obsługują następujące języki (patrz Tabela 5):

Tabela 5. Obsługiwane języki

Język	Identyfikacja języka	Pobieranie informacji	Streszczanie	Klasyfikowanie	Klastrowanie
angielski	x	x	x	x	x
niemiecki	x		x	x	x
francuski	x		x	x	x
duński	x				
fiński	x				
włoski	x		x	x	x
norweski	x				
portugalski	x		x	x	x
hiszpański	x		x	x	x
szwedzki	x				
koreański	x		x	x	x
japoński	x	x	x	x	x
chiński (tradycyjny i uproszczony)	x		x	x	x

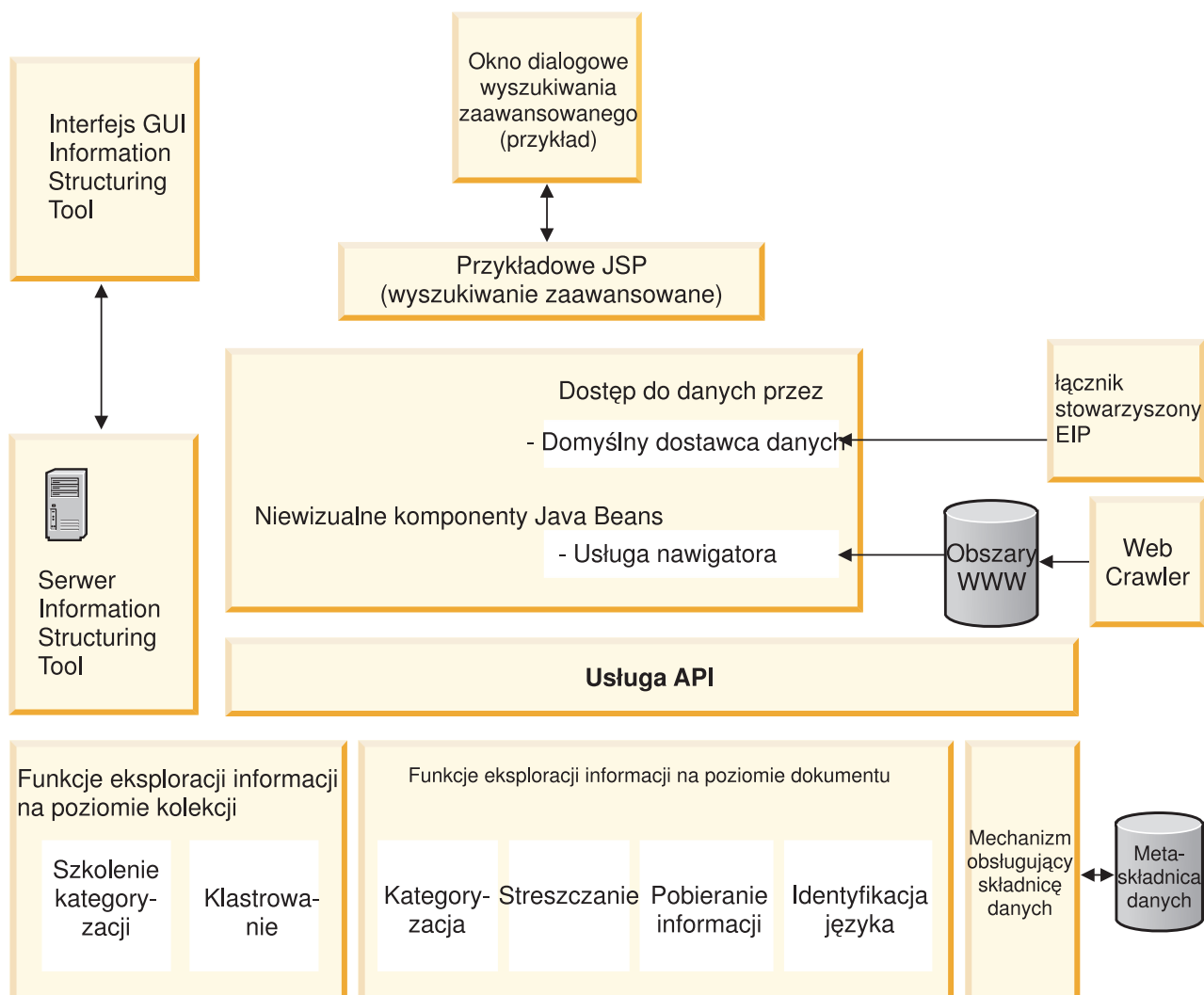
Rozdział 9, “Formaty dokumentów”, na stronie 121 zawiera listę obsługiwanych formatów dokumentów.

Koncepcja

Ilość przetwarzanych informacji wciąż rośnie. Większość organizacji posiada dużą i stale powiększającą się liczbę dokumentów elektronicznych zawierających informacje o wielkiej potencjalnej wartości, na przykład ankiety klientów - informacje strategiczne pomagające wygrać w walce rynkowej lub informacje na temat nowych kierunków rozwoju. Usługi eksploracji informacji zostały zaprojektowane jako aplikacja obsługująca wielką liczbę dokumentów elektronicznych.

Architektura systemu

Rys. 5 przedstawia architekturę systemu eksploracji informacji.



Rysunek 5. Architektura systemu eksploracji informacji

Prostokąty w prawej części diagramu odnoszą się do składników używanych przez usługi Eksploracja informacji, ale nie należy ich utożsamiać z konkretnymi produktami:

- Łącznik stowarzyszony (część EIP OO API)
- Web Crawler

W produkcie Eksploracja informacji istnieją różne warstwy:

1. API usług Java
Ta warstwa udostępnia funkcjonalność programu Eksploracja informacji i trwałość metadanych jako spójny interfejs API dla języka Java.
2. Niewizualne komponenty JavaBeans
Ta warstwa używa gotowych do zastosowania komponentów zgodnych ze specyfikacją JavaBeans.
3. Przykładowe strony JSP (Java Server Pages)

Ten poziom składa się z przykładowego kodu, używającego niewizualnych komponentów JavaBeans, które ilustrują zastosowanie zaawansowanego wyszukiwania (wyszukiwania tekstowego z ograniczeniami co do kategorii).

4. Information Structuring Tool

Aplikacja z graficznym interfejsem użytkownika służąca do tworzenia i zarządzania systematykami.

Koncepcje eksploracji informacji

Aby w pełni zrozumieć i efektywnie wykorzystać funkcjonalność eksploracji informacji, konieczne jest zrozumienie podstawowych koncepcji.

Eksploracja informacji udostępnia infrastrukturę pozwalającą na tworzenie i zarządzanie informacjami związanymi z pojedynczymi dokumentami lub kolekcjami dokumentów. Te informacje na temat dokumentu nazywa się **metadanymi**.

Biblioteka jest konceptualnym widokiem bazy danych programu Eksploracja informacji. Biblioteka składa się ze zbioru katalogów.

Katalog jest składnicą metadanych dla dokumentów tekstowych i zawiera:

- **schemat katalogu** definiujący atrybuty dla każdego dokumentu;
- **systematykę**, która jest drzewem hierarchicznym struktury **kategorii**;
- **model klasyfikacji** utworzony w oparciu o wyniki szkolenia, które mogą być używane do automatycznego przypisywania kategorii do dokumentów. Model ten jest tworzony za pomocą programu **Information Structuring Tool**, a następnie tworzone są i szkolone systematyki. Model jest używany jako dane wejściowe dla usługi klasyfikacji.

Schemat określa nazwy i typy atrybutów, które mogą być tworzone lub zapisywane dla dokumentów w katalogu. Schemat jest wstępnie zdefiniowany i zawiera następujące atrybuty:

- IKF_CONTENT typu tekstowego (string)
- IKF_TITLE typu tekstowego (string)
- IKF_AUTHOR typu tekstowego (string)
- IKF_CATEGORIES typu tekstowego (string)
- IKF_SUMMARY typu tekstowego (string)
- IKF_LANGUAGE typu tekstowego (string)
- IKF_FEATURES typu tekstowego (string)
- IKF_COMMENTS typu tekstowego (string)
- IKF_DATE typu datownik
- IKF_IDNUMBER typu liczba całkowita (integer)

Katalog tworzy **rekord** zgodnie ze schematem katalogu w celu zapisania informacji uzyskanych lub utworzonych na podstawie importowanego dokumentu. Rekord posiada niepowtarzalny identyfikator oraz zbiór par nazwa-wartość. Identyfikator ten, nazywany identyfikatorem PID, łączy tworzone rekordy z oryginalnym źródłem dokumentu.

Rys. 6 na stronie 54 przedstawia przykładowy rekord.

Rekord	
IKF_TITLE	"Ptaki"
IKF_AUTHOR	"J. Smith"
IKF_SUMMARY	"Streszczenie książki p.t. Ptaki"
IKF_CATEGORIES	Ptaki/owadożerne
IKF_DATE	07/01/2001

Rysunek 6. Przykładowy rekord

Po ustawieniu wartości dla rekordu eksploracji informacji przy użyciu komponentów JavaBean lub interfejsu API, należy sprawdzić, czy wartości mieszczą się w zdefiniowanych limitach. W przeciwnym wypadku zostanie zgłoszony wyjątek `DKIKFSizeOutOfBoundsException`. Zdefiniowane wartości to:

Tabela 6.

Klucz	Maksymalna wielkość w bajtach
IKF_CONTENT	209715
IKF_TITLE	2048
IKF_AUTHOR	2048
IKF_CATEGORIES	8192
IKF_SUMMARY	8192
IKF_LANGUAGE	8
IKF_FEATURES	524288
IKF_COMMENTS	8192

Po utworzeniu rekordu jest on przechowywany w katalogu przez przypisanie go do kategorii. Zwykle do wybrania odpowiedniej kategorii mogą być użyte wyniki klasyfikowania, można również określić kategorię na podstawie innych wartości zapisanych w rekordzie. Rekordy muszą być przypisane do kategorii, dotyczy to również indeksowania zawartości dokumentów w celu umożliwienia wykonywania wyszukiwania tekstowego. Każdy katalog ma pojedynczy indeks tekstowy, dzięki któremu wszystkie wyniki wyszukiwania są zawsze automatycznie określone w zasięgu wyszukiwania katalogu.

Mechanizm składnicy danych jest składnikiem, który zarządza dostępem do składnicy danych.

Narzędzia programu Eksploracja informacji

Usługi Eksploracja informacji udostępniają funkcje obsługujące dokumenty elektroniczne. Są to m.in.:

- program Information Structuring Tool do tworzenia i zarządzania katalogami;
- usługa identyfikacji języka, która automatycznie określa wersję językową dokumentu;
- usługa klasyfikowania, która automatycznie przypisuje dokumenty do kategorii, poprzednio zdefiniowanych za pomocą programu Information Structuring Tool;
- usługa streszczania, która analizuje słowa i zdania w dokumencie w celu jego streszczenia;
- usługa uzyskiwania informacji, która automatycznie rozpoznaje istotne elementy w tekście (bez konieczności definiowania słowników zależnych od domeny);

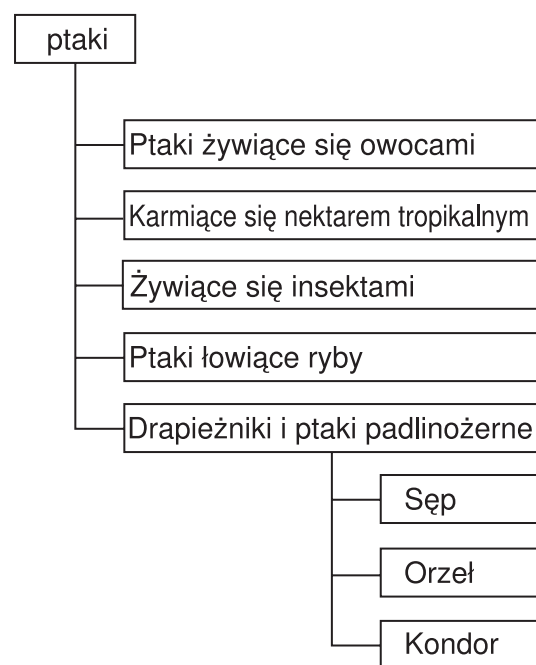
- usługa łączenia w klastry, która pozwala na utworzenie z jednego zbioru dokumentów wielu grup, zawierających podobne dokumenty (dokumenty w każdej grupie mają wspólne cechy). Klastry nie są wstępnie zdefiniowane, są tworzone automatycznie;
- usługa zaawansowanego wyszukiwania, która służy do wyszukiwania tekstu w dokumentach zapisanych w katalogu z ograniczeniem do określonych kategorii.

Program Information Structuring Tool

Narzędzie Information Structuring Tool jest aplikacją wykorzystującą interfejs WWW, która umożliwia tworzenie i zarządzanie zbiorów katalogów, zwanych bibliotekami. Katalog służy do przechowywania metadanych i jest przypisany do systematyki. Systematyka to hierarchiczna struktura kategorii służąca do klasyfikowania dokumentów zgodnie z tematyką ich zawartości.

Na przykład używając programu Information Structuring Tool, bibliotekarz może zdefiniować katalog, który ilustruje zwyczaje żywieniowe ptaków.

Rys. 7 przedstawia przykładowy katalog.



Rysunek 7. Przykładowy katalog

System, który jest dobrze podzielony na kategorie, znacznie ułatwia odnalezienie potrzebnych informacji w morzu danych. Kategorie są określane na podstawie zamierzonego użycia kolekcji dokumentów i przed użyciem są przeszkolone za pomocą przykładowych dokumentów. Model klasyfikujący, utworzony za pomocą programu Information Structuring Tool, może być używany przez usługę klasyfikującą do automatycznego przypisania kategorii do dokumentów.

Program Information Structuring Tool obsługuje następujące funkcje:

- tworzenie, zmiana nazwy i usuwanie katalogów;
- edycja opisu katalogu;
- tworzenie, zmiana nazwy i usuwanie kategorii;
- dodawanie i usuwanie dokumentów szkoleniowych z kategorii;
- przeglądanie zawartości dokumentów szkoleniowych;

- uruchamianie i zatrzymywanie procesu szkolenia katalogu;
- otrzymywanie potwierdzenia jakości szkolonych danych w katalogu.

Szczegółowe informacje na temat instalowania i używania programu Information Structuring Tool znajdują się w sekcji “Tworzenie systematyki” na stronie 63.

Identyfikacja języka

Usługa identyfikacji języka wybiera (dla określonego zbioru języków) najbardziej prawdopodobną wersję językową dokumentu.

Usługa ta zwraca ocenioną listę wersji językowych wraz z wartościami zgodności. Skróć wersji językowej jest dwuznakowym kodem kraju zgodnym z normą ISO 639. Wartości zgodności są miarą wskazującą zgodność dokumentu z danym językiem. Jest to liczba zmiennopozycyjna z zakresu od 0 (brak zgodności) do 1 (pełna zgodność). Algorytm określania wersji językowej jest zaprojektowany do rozpoznawania wersji językowych dokumentów w całości napisanych w jednym języku. Dla dokumentów wielojęzycznych nie można więc zagwarantować, że wartość zgodności będzie prawidłowa.

Wykrywane są następujące języki:

- angielski EN
- niemiecki DE
- francuski FR
- duński DA
- fiński FI
- włoski IT
- norweski (Bokmal) NB
- norweski (Nynorsk) NO
- portugalski PT
- hiszpański ES
- szwedzki SV
- koreański KO
- japoński JA
- chiński tradycyjny i uproszczony ZH

Można ustawić następujące właściwości:

- **maxResults** (tylko gdy używane jest API usług Java):
Maksymalna liczba języków, która ma być określona i zwrócona dla każdego dokumentu. Jest to liczba całkowita większa lub równa 0. Wartością domyślną jest 1, co oznacza zwrócenie najlepszego dopasowania. Jeżeli wartością jest 0, zwracane są wszystkie rozpoznawane języki w kolejności dopasowania, za wyjątkiem języków, których zgodność jest mniejsza od 0,01.

Identyfikacja języka może być użyta jako czynność wstępnego przetwarzania w innych usługach eksploracji informacji. Na przykład można użyć identyfikacji języka do wyszukania wszystkich dokumentów angielskich lub japońskich przed uzyskiwaniem informacji.

Klasyfikowanie

Klasyfikowanie oznacza przypisywanie kategorii do dokumentów i organizowanie dokumentów zgodnie z wcześniej przyjętym modelem organizacyjnym, utworzonym za pomocą programu Information Structuring Tool.

Przed użyciem usługi klasyfikowania należy zdefiniować i przeszkolić systematykę, aby utworzyć taki model za pomocą programu Information Structuring Tool.

Wynik klasyfikowania zawiera kategorię i poziom zgodności wskazujący, jak dobrze dany dokument jest dopasowany do kategorii. Dla każdego dokumentu zwracany jest zbiór takich wyników. Lista jest posortowana według wartości zgodności.

Można ustawić następujące właściwości:

- **maxResults:**
Maksymalna liczba kategorii, która może być zwrócona dla każdego dokumentu. Wartością domyślną jest -1 , co oznacza, że zwracane są wszystkie kategorie. Lista wyników jest posortowana według wartości zgodności.
- **minConfidence:**
Wartość zgodności, przypisana do dokumentu, oznacza stopień dopasowania dokumentu do kategorii. Parametr minConfidence określa minimalną wartość zgodności od 0 (brak dopasowania) do 1 (najlepsze dopasowanie). Wartością domyślną jest 0, co oznacza, że zwracane są wszystkie kategorie przypisane do dokumentu. Lista wyników jest posortowana według stopnia dopasowania.
- **catalogName:**
Określa katalog używany do klasyfikowania. Katalog może być utworzony i przeszkolony do celów klasyfikowania za pomocą programu Information Structuring Tool.

“Obsługiwane języki i formaty” na stronie 51 zawiera listę obsługiwanych wersji językowych dokumentów.

Streszczanie

Streszczanie dokumentu to zbiór zdań pochodzących z dokumentu, które są dla niego charakterystyczne. Narzędzie streszczające może na przykład pomóc w zdecydowaniu, czy dokument jest odpowiedni i czy powinien być w całości przeczytany (jest to szczególnie pomocne przy przeglądaniu listy wyników wyszukiwania).

Streszczenie jest pojedynczym łańcuchem tekstowym, który może być używany przez aplikacje jako struktura (macierz). Umożliwia to wybieranie poszczególnych zdań i określanie, czy występują one obok siebie.

Usługa streszczania może pracować w różnych trybach. Określają one sposób używania parametrów `maxLength` i `ratio` do określania długości streszczenia.

- **maxLength:**
Maksymalna liczba zdań w streszczeniu. Utworzone streszczenie nie będzie dłuższe od wartości `maxLength`. Wartość domyślna: 3.
- **ratio:**
Liczba zdań określona na podstawie długości całego dokumentu. Długość streszczenia jest określana na podstawie całkowitej długości dokumentu. Wartość domyślna: 0.1.
- **mode:**
Określa relację między parametrami `maxLength` i `ratio` niezbędną do określenia długości streszczenia. Dostępne są następujące tryby:
 - **MODE_LESS_THAN_MAXLENGTH:**
Streszczenie nie przekroczy liczby zdań określonej przez parametr `maxLength`. Jest to tryb domyślny.
 - **MODE_EQUALS_RATIO:**
Streszczenie składa się dokładnie z takiej liczby zdań, jaka jest określona przez parametr `ratio`. Liczba ta jest obliczana przez pomnożenie parametru `ratio` przez liczbę zdań w całym dokumencie.
 - **MODE_EQUALS_RATIO_BUT_AT_MOST_MAXLENGTH:**

Jak wyżej, z tą jednak różnicą, że parametr `maxLength` ogranicza maksymalną długość streszczenia.

“Obsługiwane języki i formaty” na stronie 51 zawiera listę obsługiwanych wersji językowych dokumentów.

Uzyskiwanie informacji

Ważnym zadaniem podczas analizowania dokumentów jest uzyskiwanie informacji na temat ich zawartości. Te kluczowe elementy mogą być użyte:

- do wskazania ważności informacji w celu określenia, czy dany dokument wart jest zainteresowania;
- do odszukania i zapisania kluczowych koncepcji w celu późniejszego używania w zapytaniach;
- jako kryteria zbierania dokumentów.

Przykładami kluczowych elementów są pozycje słownikowe (słowa, nazwy) lub terminy wielowyrazowe.

Dla języka angielskiego usługa uzyskiwania informacji normalizuje odszukane elementy kluczowe i grupuje ich wystąpienie w tekście, jeżeli odnoszą się do tego samego elementu lub wyrażają tę samą koncepcję. Na przykład, jeżeli w dokumencie występują nazwiska James J. Smith, Mr. Smith, James i Smith, to oznaczane są jako odnoszące się do tej samej osoby przez przypisanie ich do tej samej znormalizowanej postaci. Odmienione słowa również są przypisywane do znormalizowanej postaci, np. 'children' (dzieci) do 'child' (dziecko).

Dla języka japońskiego wszystkie elementy kluczowe są pobierane tak, jak występują w dokumencie. Żadna normalizacja nie jest wykonywana, za wyjątkiem daty, godziny i waluty, jeżeli ich postać jest zgodna z normami ISO8601 i ISO4217.

Usługa uzyskiwania informacji pozwala na analizowanie dokumentów pod kątem:

- pojedynczych i wielowyrazowych pozycji słownikowych, np. ogłoszenie, wartość, cykl produkcji;
- nazwisk osób, nazw miejsc i organizacji, np. Nowy Jork, Bush, IBM;
- skrótów, np. MB;
- dat, waluty i liczb, np. 11 stycznia 1958, 01/11/58, \$30, trzydzieści dolarów, 4,5, 5000.

Można określać trzy rodzaje informacji do uzyskiwania:

- Nazwa
- Termin
- Wyrażenie

Używając API usług Java, można identyfikować podtypy wymienione poniżej i poziomy zgodności, zaczynając od tego, jak dobrze dany podtyp pasuje do uzyskanej pozycji słownikowej. Wartością poziomu zgodności jest liczba z zakresu od 0 (źle) do 1 (dobrze). Podtypami dla nazw, terminów i wyrażen są:

- **Nazwy**
 - miejsca, np. Montreal, Londyn;
 - osoby, np. Kate Bush;
 - organizacje, np. Ministerstwo Śmiesznych Kroków;
 - nieznanne, np. Smashing Pumpkins, Krzemowa Dolina, CCTV (skrót bez pełnej postaci);
 - inne, np. Plan AIS, Konferencja ISO, Internet.

- **Terminy**
 - nieokreślone terminy, np. konglomerat rozrywkowy, świat sztuki, kod źródłowy, definicja danych, inicjatywa ulepszania procesu.
- **Wyrażenia**
 - liczebniki główne, np. cztery, pięćdziesiąt, 70;
 - liczebniki porządkowe, np. czwarty, pięćdziesiąty;
 - procenty, np. 12%, dziesięć procent;
 - data, np. 07/28/98;
 - godzina, np. czwarta dwadzieścia;
 - waluta, np. DM90, 100 złotych;
 - skróty, np. NY.

Uzyskiwanie informacji działa tylko dla dokumentów napisanych w języku angielskim i japońskim. Dla dokumentów w innych językach można użyć usługi identyfikacji języka jako wstępnej czynności przygotowującej kolekcję dokumentów. Usługa uzyskiwania informacji może być łączona z innymi funkcjami eksploracji informacji, np. może być używana jako wstępna czynność przygotowująca dla usługi streszczania, aby streszczać tylko te dokumenty, które dotyczą prezydenta Busha, a nie gubernatora stanu Texas.

Łączenie w klastry

Narzędzie to zmienia kolekcję dokumentów tak, że podobne dokumenty są łączone ze sobą, a dokumenty umieszczone w różnych grupach (klastrach) można odróżnić na podstawie zawartości. Klastrowanie może być więc użyte jako usługa pozwalająca na zorientowanie się w poruszonych tematach w dużej kolekcji dokumentów. Podobnie narzędzie to może być użyte do obsługi tworzenia systematyki za pomocą programu Information Structuring Tool przez klastrowanie przeszkolonych dokumentów w obszarze aplikacji. Klastrowanie jest również użyteczne do odszukania podobnych dokumentów w kolekcji, co może pomóc w zwróceniu uwagi na nowe technologie albo w odszukaniu bardzo podobnych dokumentów w celu wykonania analizy porównawczej.

Klastrowanie jest procesem interaktywnym, który organizuje dokumenty w klastry. Dokumenty w każdym klastrze są podobne do siebie tak bardzo, jak tylko jest to możliwe w odniesieniu do ich zawartości. Same klastry są od siebie tak różne, jak to jest możliwe. Klastrowanie działa na całej kolekcji dokumentów w przeciwieństwie do opisanych wcześniej usług eksploracji informacji, takich jak klasyfikowanie i streszczanie, które działają na poziomie dokumentu. Klastrowanie działa przez porównywanie odpowiednich cech dokumentu z innymi i grupowaniu ich na podstawie podobieństw.

Podczas fazy klastrowania do zbioru dokumentów nie można dodawać nowych dokumentów.

Można ustawić następujące właściwości:

- **maxClusterCount**
Jest to maksymalna liczba tworzonych klastrow.
- **minClusterCount**
Jest to minimalna liczba tworzonych klastrow.
- **clusterFeatureCount**
Liczba etykiet (słów kluczowych), które są zwracane dla każdego klastra.

Wartości te nie są powiązane z klastrem, ale są wytycznymi dla całego procesu. Usługa klastrowania zwraca listę wyników.

Klastrowanie działa tylko dla języka angielskiego. Dla dokumentów w innych językach można użyć usługi identyfikacji języka jako wstępnej czynności przygotowującej kolekcję dokumentów.

Wyszukiwanie zaawansowane

W porównaniu do standardowego wyszukiwania EIP, które jest obsługiwane przez cały serwer danych Enterprise Information Portal, wyszukiwanie zaawansowane jest dostępne tylko dla dokumentów, których identyfikatory są zapisane w katalogu utworzonym przez program Information Structuring Tool. W celu dalszego zawężenia wyszukiwania, wyszukiwanie zaawansowane nie tylko wyszukuje tekst, ale również ogranicza wyszukiwanie tylko do jednej kategorii.

Można ustawić następujące parametry:

- **catalogName:**
Określa katalog używany do wyszukiwania. Katalog może być tworzony i szkolony za pomocą programu Information Structuring Tool.
- **maxResults:**
Jest to maksymalna liczba wyników zwracanych w jednym zapytaniu. Wartością domyślną jest 0, co oznacza brak limitu (zwracane są wszystkie wyniki).

Można używać następujących zapytań:

1. Zapytanie czysto tekstowe. To zapytanie zwraca wszystkie dokumenty, które pasują do zapytania. Lista wyników jest posortowana według wartości zgodności.
2. Zapytanie dotyczące tylko kategorii. To zapytanie zwraca wszystkie dokumenty, które są przypisane do danej kategorii. Wyniki są zwracane w nieokreślonej kolejności.
3. Zapytanie łączone. W tym zapytaniu zwracane są wszystkie dokumenty, które pasują do zapytania i należą do określonej kategorii. Lista wyników jest posortowana według wartości zgodności.

Zapytania zaawansowane są zawsze wykonywane w określonym katalogu. Określa się to jako zasięg wyszukiwania katalogu. Wyszukiwanie w wielu katalogach jednocześnie nie jest możliwe, ponieważ katalog reprezentuje widok importowanych dokumentów, który musi być zachowany.

Składnia zapytania (BNF) jest następująca:

```
query_string      ::= term
term              ::= ( term )
                  ::= single_term
                  ::= compound_term
single_term       ::= category_term
                  ::= text_search_term
                  ::= string_term
                  ::= number_term
compound_term     ::= term binary_bool_operator term
                  ::= unary_bool_operator single_term
category_term     ::= ( DKIKFCategory category_operator category_path_value )
text_search_term  ::= ( "attribute name" CONTAINS text_search_value )
string_term       ::= ( "attribute name" string_operator string_value )
number_term       ::= ( "attribute name" basic_operator number_value )
binary_bool_operator ::= AND | OR
unary_bool_operator  ::= NOT
category_operator   ::= >= | =
string_operator     ::= LIKE | basic_operator
basic_operator      ::= > | < | <= | >= | != | =
category_path_value ::= "category path"
text_search_value   ::= "string"
string_value        ::= "string"
number_value        ::= "integer" | "decimal number"
```


- Terminy, łańcuchy tekstowe i liczby reprezentują wspólne terminy.
- Operator kategorii '=' ogranicza zasięg wyszukiwania do jednej kategorii.
- Operator kategorii '>=' powiększa zasięg wyszukiwania do podanej kategorii i wszystkich podkategorii w drzewie kategorii.
- Wyszukiwanie tekstowe z klauzulą **CONTAINS** może zawierać znaki zastępcze ('_') dla jednego znaku i ('%') dla ciągu znaków. Np. **_LOB** może oznaczać **BLOB** i **CLOB**, a **nazwa%** może oznaczać **nazwa pliku**. Tylko atrybuty schematów, które zostały oznaczone jako dostępne w wyszukiwaniu tekstowym, np. **IKF_CONTENT**, mogą być użyte z operatorem **CONTAINS**.
- Tekst wyszukiwania w klauzuli **LIKE** może również zawierać znaki zastępcze, używane w języku SQL.
- Pełna lista obsługiwanych nazw atrybutów znajduje się w sekcji "Koncepcje eksploracji informacji" na stronie 53.

Przykłady:

- Zapytanie czysto tekstowe:

```
("IKF_CONTENT" CONTAINS "'Afryka Południowa'") AND NOT
("IKF_CONTENT" CONTAINS "'Przylądek'")
```
 - Zapytanie dotyczące kategorii:

```
("DKIKFCATEGORY" >= "ptaki/roślinożerne")
```
 - Zapytanie łączone:

```
("IKF_CONTENT" CONTAINS "'Afryka Południowa'") AND
("DKIKFCATEGORY" >= "ptaki/drapieżne i padlinożercy/Sokół")
```
 - Zapytanie dotyczące atrybutów:

```
("IKF_SUMMARY" LIKE "ptaki lasów tropikalnych")
```
- lub
- ```
("IKF_FEATURES" LIKE "Goethe") AND ("IKF_TITLE" = "Faust")
```

## Interfejs programisty

Funkcjonalność eksploracji informacji dostępna jest również do tworzenia aplikacji z wykorzystaniem:

- API usług Java
- Komponentów JavaBean

**API usług Java** integruje pełną funkcjonalność eksploracji informacji, z wyjątkiem konserwacji katalogów, która jest częścią produktu Information Structuring Tool jako usługi Enterprise Information Portal. Udostępnia on komunikację typu klient-serwer w oparciu o Java RMI.

Aplikacje używające API usług Java mogą:

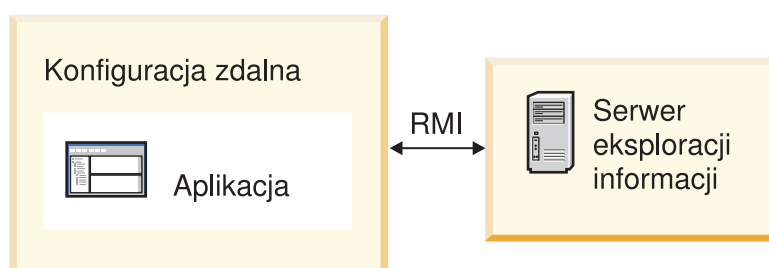
- określać, w jakiej wersji językowej jest napisany dany dokument;
- tworzyć streszczenia dokumentów tekstowych;
- przypisywać kategorie do dokumentów;
- uzyskiwać informacje z dokumentu tekstowego, np. nazwy, terminy, wyrażenia;
- łączyć w grupy podobne dokumenty;
- zapisywać i wyszukiwać metadane dokumentów w katalogu;
- wykonywać wyszukiwanie tekstowe w dokumentach, ograniczone do określonych kategorii i atrybutów, np. streszczenia.

API usług Java może być uruchomione w trybie lokalnym, używając bezpośredniego wywołania metody, lub zdalnym, za pomocą RMI (Java Remote Method Invocation). Tryb zdalny umożliwia skonfigurowanie jednego serwera jako serwera eksploracji informacji, który wykonuje analizę tekstową, indeksowanie i wyszukiwanie. Mechanizm zadań serwera obejmuje wysyłanie kompletnych zadań do serwera eksploracji informacji (zdalnego komputera) wraz z całym przetwarzaniem wykonywanym na tym serwerze.

Gdy używany jest program Web Crawler, odpowiedni mechanizm dostępowy musi być zaimplementowany za pomocą komponentów JavaBeans. Dostęp do programu Web Crawler nie jest możliwy na poziomie API usług Java.

Dokładny opis API usług eksploracji informacji zawiera publikacja *Application Programming Guide for Windows*.

Rys. 8 na stronie 62 przedstawia zdalną konfigurację eksploracji informacji.



Rysunek 8. Zdalna konfiguracja eksploracji informacji

**Komponenty eksploracji informacji** to interfejs API wysokiego poziomu języka Java używane do szybkiego tworzenia aplikacji. Są one zgodne ze specyfikacją JavaBeans. Komponenty te nie obsługują serwerowego przetwarzania zadań, tak więc tworzenie aplikacji z wykorzystaniem komponentów JavaBeans powinno być wykonywane na tym samym komputerze.

Każdy z komponentów może być użyty w formie obsługującej zdarzenia, niektóre mogą być wywołane bezpośrednio. Integracja z istniejącymi komponentami Enterprise Information Portal jest wykonywana w oparciu o obsługę zdarzeń wynikowych, używaną przez komponenty EIP. Oznacza to, że wyszukiwanie stowarzyszone i zdarzenia wynikowe programu Web Crawler są kompatybilne z usługami eksploracji informacji, dlatego wyniki eksploracji informacji mogą być przetworzone przez program EIP. Szczegółowy opis komponentów eksploracji informacji zawiera publikacja *Application Programming Guide for Windows*.

---

## Pierwsze kroki

Przewodnik Pierwsze kroki prezentuje inżynierom wiedzy, administratorom lub programistom, jak można wykorzystać technologię eksploracji informacji firmy IBM w rzeczywistych zastosowaniach biznesowych. Przewodnik jest podzielony w następujący sposób:

- Krótkie wprowadzenie
- Czynności, które muszą być wykonane przed uruchomieniem programu Pierwsze kroki
- Organizowanie danych przykładowych
- Dostęp do danych
- Używanie klienta przykładowego
- Usuwanie danych przykładowych

- Literatura

Aby uzyskać dostęp do aplikacji EIP eksploracja informacji - pierwsze kroki, należy otworzyć plik <CMBROOT>\ikf\firststeps\first\_steps.html.

Aby uruchomić przewodnik Eksploracja informacji - Pierwsze kroki w systemie Windows, kliknij przycisk Start i wybierz kolejno opcje **Enterprise Information Portal for Multiplatforms 8.1** → **<Eksploracja informacji>** → **Pierwsze kroki**.

Publikacja EIP eksploracja informacji - pierwsze kroki działa również jako program weryfikujący instalację.

---

## Tworzenie systematyki

Narzędzie Information Structuring Tool jest aplikacją wykorzystującą interfejs WWW, która umożliwia tworzenie i zarządzanie zbiorem katalogów, noszącym nazwę biblioteki. Katalog służy do przechowywania metadanych i jest przypisany do systematyki, która organizuje przechowywane w nim dane. Systematyka to hierarchiczna struktura kategorii służąca do klasyfikowania dokumentów zgodnie z tematyką ich zawartości. Na przykład najwyższy poziom systematyki może zawierać kategorie, takie jak *biznes*, *kultura*, *sport*, a *sport* z kolei podkategorie *gry zespołowe* i *lekkoatletyka*. Na kolejnym niższym poziomie *gry zespołowe* mogą dzielić się na *piłkę nożną*, *baseball*, *tenis*.

Przypisując do kategorii dokumenty szkoleniowe, narzędzie Information Structuring Tool tworzy model klasyfikacji, który może następnie być wykorzystany przez składniki bean usługi klasyfikującej w celu przypisania dokumentowi kategorii.

## Instalowanie programu Information Structuring Tool

Program Information Structuring Tool musi być zainstalowany jako aplikacja internetowa w pojemniku serwletu, np. w mechanizmie obsługi serwletów IBM WebSphere Application Server (WAS).

Nie jest jednak dozwolone jednoczesne używanie dwóch aplikacji Information Structuring Tool (np. o nazwach IST1 i IST2), korzystających z tej samej instancji Eksploracji informacji.

Przed instalacją programu Information Structuring Tool należy sprawdzić, czy została zainstalowana i skonfigurowana poprawna wersja serwera WAS. Więcej informacji na ten temat zawiera publikacja *Planowanie i instalowanie programu Enterprise Information Portal*.

Do zainstalowania programu Information Structuring Tool wymagane są następujące prawa dostępu użytkownika:

- W systemie Windows: uprawnienia administratora
- W systemie AIX: uprawnienia użytkownika root
- W systemie Solaris: uprawnienia użytkownika root

## Pierwsze kroki

Aplikacja udostępnia interfejs WWW pozwalający na definiowanie, zarządzanie i szkolenie systematyk. Jest on podzielony na dwie ramki. Ramka znajdująca się po lewej stronie nazywana jest widokiem *katalogu* i służy do tworzenia i zarządzania systematykami. Ramka po prawej stronie nosi nazwę *notesu* i prezentuje informacje na temat korzystania z aplikacji. Notes wyświetla serię kart służących do szkolenia i sprawdzania dokumentów. Aby można było korzystać z aplikacji, należy najpierw utworzyć w lewej ramce katalogi i kategorie.

## Prawa dostępu

Przy korzystaniu z narzędzia Information Structuring Tool informacjami o nazwie użytkownika i jego hasle zarządza produkt Enterprise Information Portal. Użytkownik podaje nazwę konta i hasło, które zna tylko Enterprise Information Portal.

Po uruchomieniu programu Information Structuring Tool wyświetlony zostanie komunikat ostrzegający, że aplet Java, używany do wysyłania dokumentów szkoleniowych, wymaga praw do odczytywania plików w systemie. Jeżeli użytkownik odmówi przyznania tych praw, program nie będzie mógł wysłać dokumentów szkoleniowych i w efekcie szkolenie systematyki nie będzie możliwe.

Narzędzie Information Structuring Tool może być używane w środowisku wieloużytkownikowym. Aby była możliwa jednoczesna praca wielu użytkowników nad jedną systematyką, narzędzie Information Structuring Tool obsługuje mechanizm blokowania, służący do kontrolowania dostępu do katalogu i jego kategorii.

Użytkownik może zastosować jawną blokadę katalogu, zaznaczając go i blokując przed rozpoczęciem pracy. Może również rozpocząć pracę z katalogiem, czyli zacząć dodawać lub zmieniać nazwy kategorii i trenować dokumenty. W takim wypadku katalog zostanie automatycznie zablokowany w celu uniknięcia konfliktów dostępu. Inni użytkownicy mogą wyświetlać ten katalog, ale nie mogą w nim niczego zmieniać do chwili odblokowania go przez użytkownika, który ten katalog zablokował.

Należy zauważyć, że jeśli serwer, na którym jest zainstalowane narzędzie Information Structuring Tool, zostanie zrestartowany, wszystkie blokady zostaną usunięte.

## Definiowanie systematyki

Katalog jest punktem wyjścia dla systematyki, która jest strukturą drzewiastą złożoną z kategorii.

Aby zdefiniować nowy katalog i wybrać odpowiednie kategorie, należy wykonać następujące czynności:

1. Zdecydować, jakie kategorie należy zdefiniować oraz utworzyć nową systematykę.

W widoku katalogów wybierz opcję **Biblioteka** i kliknij ją prawym przyciskiem myszy. Zostanie wyświetlone menu. Po wybraniu pozycji **Nowy katalog** zostanie utworzona ikona katalogu. Zmień nazwę tej ikony, podając nazwę katalogu i kliknij przycisk **Enter**. Zostanie utworzony folder o tej samej nazwie. Jest to kategoria podstawowa. Zawartość notesu ulegnie zmianie.

Inną alternatywną metodą dodania nowego katalogu do biblioteki jest zaimportowanie istniejącej systematyki (utworzonej poza programem Information Structuring Tool, np. w systemie plików). Więcej informacji zawiera sekcja “Wysyłanie dokumentów szkoleniowych” na stronie 66.

Aby aktywnie pracować z systematyką, należy zablokować katalog. Operacja ta jest wykonywana automatycznie podczas tworzenia katalogu. Aby zablokować istniejący katalog, wybierz go i kliknij prawym przyciskiem myszy. Zostanie wyświetlone menu. Wybierz pozycję **Zablokuj katalog**. Ikona katalogu ulegnie zmianie. Ikony służą do wyświetlania różnych typów statusu katalogu:



Drzewo systematyki jest zwinięte i nie jest zablokowane przez żadnego użytkownika.



Drzewo systematyki jest rozwinięte i nie jest zablokowane przez żadnego użytkownika.



Drzewo systematyki jest zwinięte i jest zablokowane przez bieżącego użytkownika.



Drzewo systematyki jest rozwinięte i zablokowane przez bieżącego użytkownika.

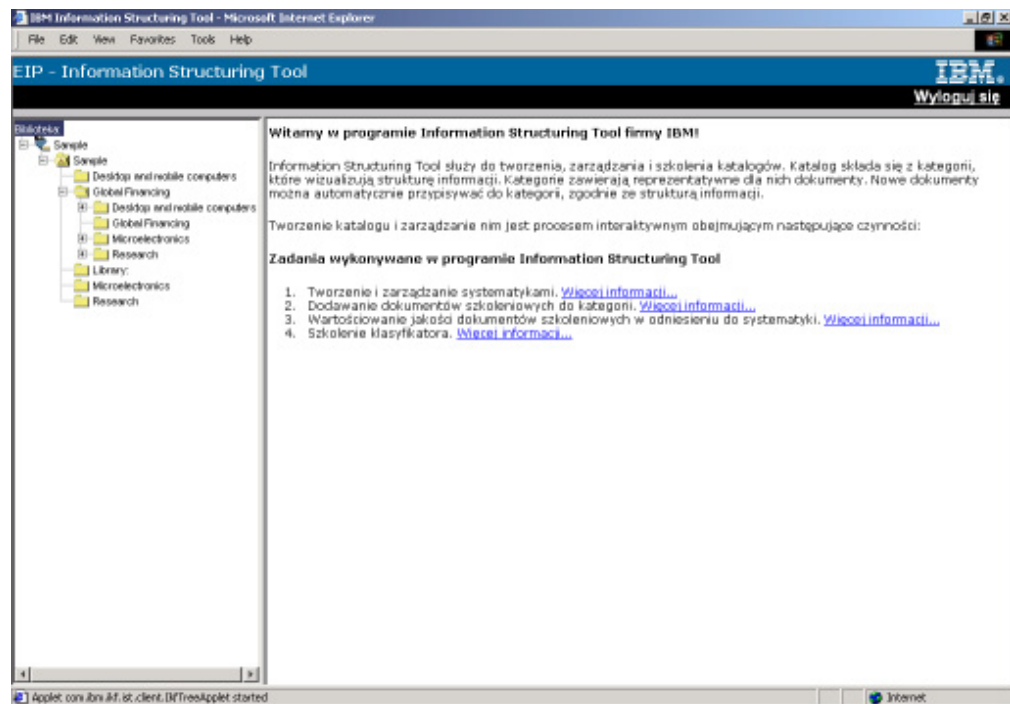


Drzewo systematyki jest zwinięte i zablokowane przez innego użytkownika.



Drzewo systematyki jest rozwinięte i zablokowane przez innego użytkownika.

Rys. 9 przedstawia przykład dwóch katalogów.



Rysunek 9. Przykład katalogu

- 
2. Po utworzeniu i zablokowaniu katalogu można do niego dodawać i usuwać kategorie oraz zmieniać ich nazwy. Po utworzeniu nowego katalogu tworzona jest także kategoria podstawowa o takiej samej nazwie. Można zmienić nazwę tej kategorii. W tym celu wybierz kategorię i kliknij ją prawym przyciskiem myszy. Wybierz z menu pozycję **Zmień nazwę**, wpisz nową nazwę i naciśnij **Enter**.

Aby dodać nowe kategorie, zaznacz kategorię, do której chcesz dodać podkategorie i kliknij ją prawym przyciskiem myszy. Wybierz z menu pozycję **Nowa kategoria**, wpisz nazwę nowej kategorii i naciśnij **Enter**. Nazwy kategorii znajdujących się na tym samym poziomie drzewa muszą być unikalne.

Kategorii podstawowej nie można usunąć. Jest ona usuwana tylko wraz z całym katalogiem. Usunięcie kategorii spowoduje również usunięcie wszystkich podkategorii (kategorii niższego rzędu w strukturze drzewa), które mogą zawierać dokumenty szkoleniowe i wszystkie rekordy.

3. Następne informacje dotyczące wybranego katalogu są wyświetlane na zakładce **Właściwości**. Za pierwszym razem pole opisu jest puste. Aby dodać lub edytować opis katalogu, kliknij przycisk **Edytuj opis**. Zostanie wyświetlone okno, w którym można wpisać opis.

## Wybieranie dokumentów szkoleniowych

Jakość modelu klasyfikowania jest zależna od jakości przypisanych do każdej kategorii dokumentów szkoleniowych.

Muszą to być dokumenty w jednym z formatów obsługiwanych przez produkt Enterprise Information Portal. Rozdział 9, "Formaty dokumentów", na stronie 121 zawiera listę obsługiwanych formatów.

Bardzo istotne jest wybranie odpowiedniego zbioru dokumentów szkoleniowych. Wybrane dokumenty powinny:

- Być reprezentatywne dla kategorii.
- Zawierać odpowiednią ilość opisu tekstowego bez zbyt wielkiej liczby znaczników lub list słów.
- Wszystkie powinny być napisane w takim samym stylu, np. należy unikać mieszania dokumentów napisanych językiem potocznym i oficjalnym.
- Mieć zbliżoną i niezbyt wielką długość; dokumenty szkoleniowe ponadto powinny mieć długość zbliżoną do dokumentów, które mają podlegać kategoryzowaniu przy użyciu odpowiedniej usługi.

Zaleca się użycie zbioru około 40 dokumentów szkoleniowych dla każdej kategorii. Jeśli jednak wybrana kategoria jest bardzo ogólna, wymagane jest użycie większej liczby dokumentów. Kategorie muszą być starannie dobrane i powinny być znaczące. Kategorie i dokumenty szkoleniowe, które dla osoby indeksującej są niejednoznaczne i nieprzejryste, na pewno będą sprawiać problemy podczas automatycznego przetwarzania.

## Wysyłanie dokumentów szkoleniowych

Aby dodać dokumenty szkoleniowe, należy wybrać odpowiednią kategorię i wyświetlić okno **Lista dokumentów szkoleniowych**. Aby dodać do tej listy dokument, kliknij przycisk **Dodaj dokument...**

Wyświetlone zostanie okno **Dodaj dokumenty szkoleniowe**. Po zakończeniu przesyłania plików nie ma konieczności zamykania tego okna, może być użyte do przesłania plików do innej kategorii lub katalogu.

Aby dodać dokumenty szkoleniowe, kliknij przycisk **Przeglądaj...** i w oknie **Otwórz** wybierz odpowiednie pliki.

Można wybrać do przesłania jeden lub kilka plików z jednego katalogu albo cały katalog. Jeżeli wybrany katalog jest pusty, wyświetlony zostanie odpowiedni komunikat.



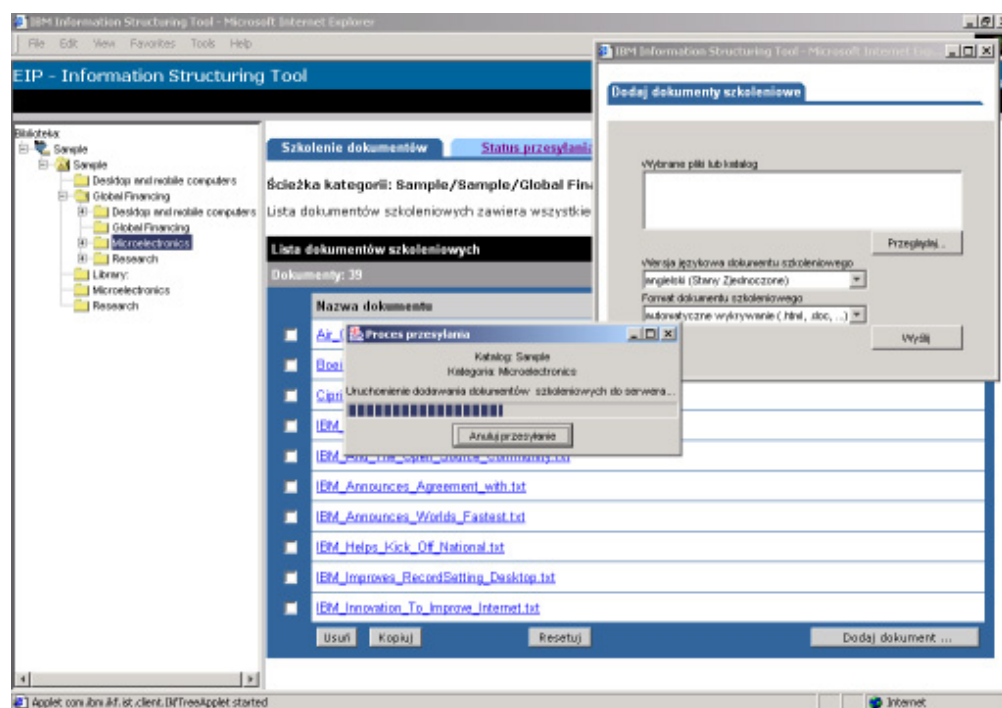
Pozwala to na zaimportowanie i pracę z istniejącymi systematykami, które zostały utworzone poza programem Information Structuring Tool, na przykład w systemie plików.

Jeżeli wybrany katalog w systemie plików (np. aplikacje) zawiera podkatalog projekt, kategoria aplikacji w drzewie systematyki również będzie posiadać podkatalog projekt. Pliki z tego podkatalogu są dodawane do podkategorii. Jeżeli podkategoria nie istnieje, zostanie utworzona, a następnie zostaną do niej skopiowane pliki.

Wybierz język dokumentu i sformatuj wszystkie wybrane pliki. Zawsze używaj formatu "automatyczne wykrywanie", z wyjątkiem plików tekstowych.

Aby dodać dokument do listy dokumentów szkoleniowych, kliknij przycisk **Wyślij**.

Rys. 10 przedstawia dodawanie dokumentów szkoleniowych.



Rysunek 10. Dodawanie dokumentów szkoleniowych

Wyświetlone zostanie okno **Proces wysyłania**. Istnieje możliwość anulowania operacji przysyłania w tym oknie. Po anulowaniu tej operacji kolejne pliki nie są dodawane. Pliki, które zostały dodane jako dokumenty szkoleniowe, będą usunięte. Jednak, jeżeli podczas przysyłania dodane zostały podkategorie, po anulowaniu nie będą one usunięte.

Jeżeli pliki nie mogą być przesłane pomyślnie, w oknie **Status wysyłania** zostanie wyświetlona przyczyna niepowodzenia. Na przykład:

- Istnieje już plik o takiej samej nazwie.
- Plik jest pusty.
- Plik nie mógł być przesłany do serwera.

Aby powrócić do okna Dokumenty szkoleniowe, należy kliknąć kartę **Dokumenty szkoleniowe**. Wszystkie dokumenty szkoleniowe, które zostały pomyślnie przesłane do kategorii, są w tym oknie wyświetlane. Aby zobaczyć wszystkie dokumenty na liście, należy użyć przycisków **Wstecz** i **Dalej**.

Jeżeli te same dokumenty mają być dodane jako dokumenty szkoleniowe do kilku kategorii, należy przesłać je do jednej kategorii, a następnie skopiować do innych kategorii. Nie ma potrzeby ponownego przesyłania plików. Aby skopiować pliki, należy zaznaczyć jeden lub kilka dokumentów na liście i kliknąć przycisk **Kopiuuj**. W wyświetlanym oknie kliknij przycisk **Przeglądaj...**, aby wybrać jedną lub więcej kategorii, do których mają być skopiowane dokumenty i kliknij przycisk **Wyślij**.

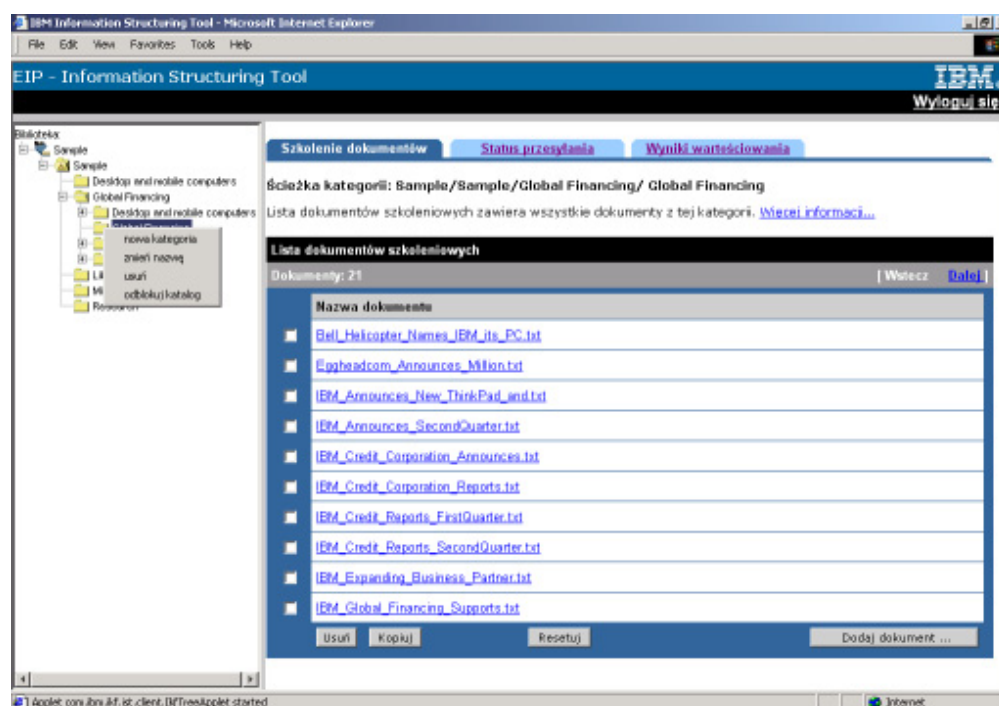
Podczas przesyłania pliku nie jest dozwolone wykonywanie następujących czynności:

- odblokowywanie katalogu;
- wylogowywanie się z programu Information Structuring Tool;
- uruchamianie szkolenia katalogu lub wartościowania;
- zmienianie nazwy katalogu;
- usuwanie statusu przesyłania z katalogu.

Możliwe jest wykonywanie następujących czynności:

- uruchamianie innego procesu przesyłania do tej samej lub innej kategorii;
- praca z innym katalogiem.

Rys. 11 przedstawia listę dokumentów szkoleniowych.



Rysunek 11. Lista dokumentów szkoleniowych

Jeżeli przeglądarka zostanie zamknięta podczas przesyłania plików, pliki zostaną dodane jako dokumenty szkoleniowe, jeśli zostały w całości przesłane do serwera. W przeciwnym przypadku żadne pliki nie będą dodawane.

Jeżeli wysłane zostały wszystkie dokumenty szkoleniowe, to aby rozpocząć wartościowanie, należy wybrać katalog.



## Wartościowanie modelu klasyfikacji

Po zdefiniowaniu systematyki i przypisaniu dokumentów szkoleniowych do każdej kategorii, należy przystąpić do wartościowania systematyki. Wartościowanie systematyki pomaga w ustaleniu, jak dobrze zostały dobrane dokumenty szkoleniowe dla przyjętej systematyki. Proces ten jest interaktywny i składa się z następujących kroków:

1. uruchamianie wartościowania;
2. oszacowanie wyników;
3. zmiana w systematyce lub dokumentach szkoleniowych;
4. ponowne uruchamianie wartościowania.

Podczas każdej iteracji proces wartościowania:

- Dzieli dokumenty szkoleniowe na zbiór szkoleniowy (ok. 80% dokumentów) i zbiór testowy (ok. 20% dokumentów). Następnie szkolony jest katalog z wykorzystaniem zbioru szkoleniowego, a potem uruchamiana jest usługa klasyfikacji dla zbioru testowego.
- Sprawdza, czy dokumenty zostały przypisane do poprawnych kategorii z wystarczająco wysokim poziomem zgodności. Mieści się on w zakresie od 0 do 1, gdzie 1 jest wynikiem idealnym. Wartość zgodności może być ustawiona. Wartością domyślną jest 0,5.

Można wybrać od trzech do pięciu iteracji. Wartością domyślną są trzy iteracje, które pozwalają na poznanie silnych i słabych stron systematyki. Gdy wybranych zostanie pięć iteracji, wszystkie dokumenty będą umieszczone w zbiorze szkoleniowym i testowym.

Naciśnij **Uruchom wartościowanie**, aby uruchomić proces wartościowania.

Dla każdej kategorii w każdej iteracji procesu wartościowania obliczane są następujące parametry:

- *Poprawne* dokumenty. Liczba dokumentów szkoleniowych kategorii c przypisanych do tej kategorii podczas wartościowania.
- *Wychodzące* dokumenty. Liczba dokumentów szkoleniowych kategorii c przypisanych do innej kategorii podczas wartościowania.
- *Przychodzące* dokumenty. Liczba dokumentów szkoleniowych przypisanych do kategorii c podczas wartościowania, ale przypisanych początkowo do innej kategorii.
- *Nieprzypisane* dokumenty. Liczba dokumentów szkoleniowych kategorii c nie przypisanych do żadnej kategorii podczas wartościowania. Mogą być tutaj zliczane dokumenty, które zostały przypisane do tej kategorii, ale ich poziom zgodności jest niższy od ustalonego.

Podczas przeglądu procesu wartościowania na poziomie katalogu wyświetlane są następujące informacje:

- Status wartościowania wskazuje, czy proces wartościowania jest uruchomiony, czy zatrzymany.
- Ostatnie wartościowanie wyświetla datę poprzedniego wartościowania.
- Liczba zakończonych iteracji wartościowania.
- Całkowita liczba iteracji.
- Średnia precyzja, czyli średni procentowy udział poprawnych dokumentów, które od początku były przypisane do kategorii, jak również dokumentów przychodzących.
- Średnia poprawność, czyli średni procentowy udział poprawnych dokumentów w danej kategorii.
- Poprawne dokumenty - liczba dokumentów poprawnie przypisanych do kategorii.
- Niepoprawne dokumenty - liczba dokumentów przychodzących lub wychodzących.
- Nieprzypisane dokumenty.

Precyzja i poprawność są blisko powiązane z przypisaną wartością zgodności. Jeżeli wartość zgodności jest niska, wartość precyzji spada, a rośnie wartość poprawności i na odwrót. Wysoka wartość precyzji oznacza, że wiele dokumentów szkoleniowych zostało poprawnie przypisanych. Wysoka wartość poprawności oznacza, że większość dokumentów zostało przypisanych do kategorii (nie pozostało wiele nieprzypisanych dokumentów).

## Wyniki wartościowania

Szczegółowe wyniki wartościowania są dostępne po kliknięciu karty **Wyniki wartościowania**. Jeżeli po wybraniu katalogu kliknięta zostanie ta karta, wyświetlone zostaną wyniki dla całego katalogu, jeżeli wybrana zostanie kategoria - wyświetlone zostaną wyniki dla tej kategorii.

Rys. 12 przedstawia wyniki wartościowania na poziomie katalogu:

**Wyniki wartościowania dla każdej kategorii**

Wartościowane kategorie: 4

| Kategoria                           | Dokładność | Średnia poprawność | Przychodzące | Wychodzące | Poprawne | Nieprzypisane |
|-------------------------------------|------------|--------------------|--------------|------------|----------|---------------|
| Sample/Research                     | 100,0      | 100,0              | 0,0          | 0,0        | 100,0    | 0,0           |
| Sample/Microelectronics             | 100,0      | 97,0               | 0,0          | 2,0        | 97,0     | 0,0           |
| Sample/Desktop and mobile computers | 87,0       | 81,0               | 10,0         | 8,0        | 72,0     | 8,0           |
| Sample/Global Financing             | 85,0       | 85,0               | 12,0         | 12,0       | 75,0     | 0,0           |

Rysunek 12. Wyniki wartościowania na poziomie katalogu

Całościowe wyniki katalogu. Wartości wyświetlane w kolorze czerwonym (krytyczne) i niebieskim (mniej krytyczne) wskazują, że coś nie jest w porządku albo z kategorią, albo z dokumentami szkoleniowymi. Ponieważ jakość kategorii jest określana przez jej dokumenty szkoleniowe, należy obserwować na poziomie kategorii przenoszenie dokumentów między kategoriami.

Zmiany w kategoriach lub dokumentach szkoleniowych są silnie zależne od tego, czy nacisk został położony równo na precyzji i poprawności, czy większy nacisk jest położony na precyzję. Wyższe wartości precyzji oznaczają kategorię bardziej odróżniającą się od innych kategorii w systematyce. Z drugiej strony wyższe wartości precyzji oznaczają mniej nieprzypisanych dokumentów szkoleniowych.

Wyniki szkolenia wyświetlane są na dwóch poziomach:

### 1. Poziom katalogu:

Dla każdej kategorii w katalogu:

- Poprawność i precyzja (w procentach)

- Dokumenty przychodzące (w procentach)
- Dokumenty wychodzące (w procentach)
- Dokumenty poprawne (w procentach)
- Dokumenty nieprzypisane (w procentach)

## 2. Poziom kategorii:

Dla dokumentów przychodzących i wychodzących:

- Dokumenty szkoleniowe i kategorie początkowe lub końcowe

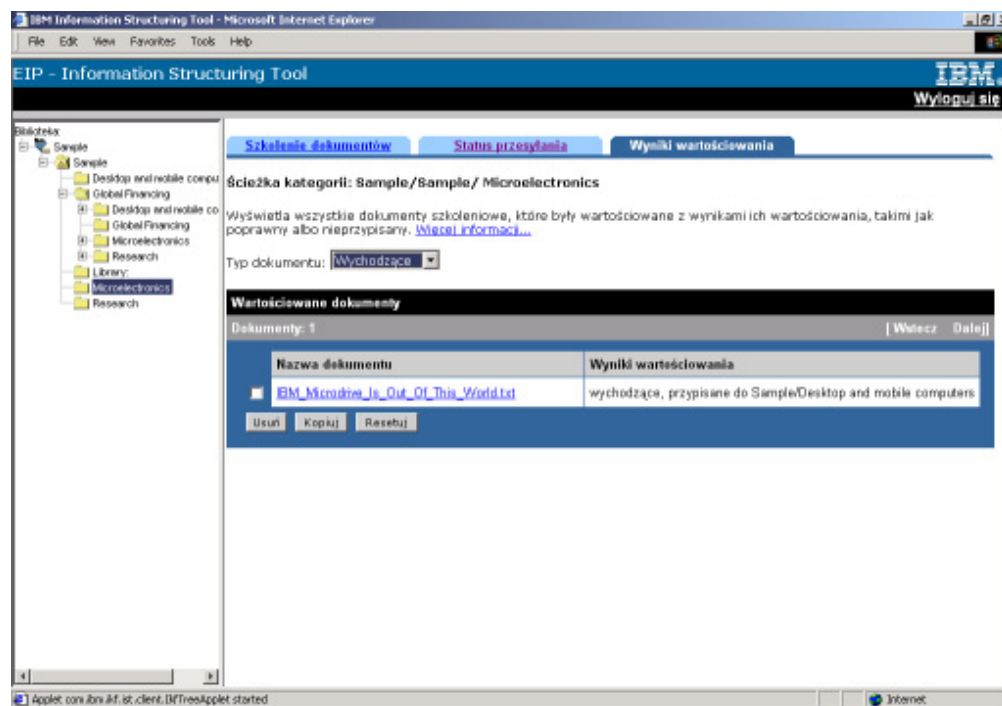
Dla dokumentów poprawnych i nieprzypisanych:

- Dokumenty szkoleniowe

## Interpretowanie wyników wartościowania

W tej sekcji omówione zostało interpretowanie wyników wartościowania. Trzeba zawsze pamiętać o tym, że systematyka jest niepodzielna, co oznacza, że zmiany w jednej sekcji systematyki mogą mieć ujemny wpływ na wyniki w innej sekcji systematyki.

Rys. 13 przedstawia wyniki wartościowania na poziomie kategorii:



Rysunek 13. Wyniki wartościowania na poziomie kategorii

Zaczynając pracę z wynikami wartościowania na poziomie katalogu, należy wybrać kategorie z potencjalnie niskimi wartościami precyzji i poprawności i z dokumentami oznaczonymi kolorem czerwonym (krytyczne) i niebieskim (mniej krytyczne). Dla każdej kategorii w tej iteracji (dostępnych do wybrania w lewej ramce), należy sprawdzić:

- Dużą liczbę dokumentów **przychodzących**:
  - Kategoria pobiera dokumenty z innych kategorii:
    - Kategoria nie jest wystarczająco wyróżniająca. Skopiuj dokumenty do tej kategorii.
    - Uczyń kategorię bardziej wyróżniającą przez wybranie bardziej odpowiednich dokumentów szkoleniowych albo podziel kategorię na podkategorie.
  - Kategoria pobiera dokumenty z jednej lub dwóch innych kategorii:

- Skopiuj dokumenty do tej kategorii.
- Sprawdź oryginalną kategorię. Czy jest sens utrzymywania tej kategorii lub czy ta kategoria powinna być połączona z inną? Aby je połączyć, skopiuj wszystkie dokumenty szkoleniowe do kategorii, którą chcesz zachować, i usuń pozostałe.
- Dużą liczbę dokumentów **wychodzących**:
  - Kategoria traci dokumenty na rzecz innych kategorii:
    - Kategoria nie jest wystarczająco wyróżniająca. Wybierz lepiej dobrane dokumenty szkoleniowe albo rozważ usunięcie tej kategorii.
  - Kategoria traci dokumenty na rzecz jednej lub dwóch innych kategorii:
    - Skopiuj dokumenty do innych kategorii.
    - Sprawdź tę kategorię. Czy jest sens utrzymywania tej kategorii lub czy ta kategoria powinna być połączona z inną?
- Dużą liczbę **nieprzypisanych** dokumentów:
  - Porównaj przypisane dokumenty z poprawnymi pod względem:
    - rozmiaru
 

Jeżeli przypisane dokumenty są krótsze, rozważ połączenie dwóch dokumentów w celu utworzenia dłuższego dokumentu, którego wielkość będzie zbliżona do wielkości poprawnych dokumentów.
    - stylu
 

Jeżeli style są różne, usuń dokument.
    - tematu
 

Jeżeli temat jest nieco inny, ale wciąż jest powiązany z tematem poprawnych dokumentów, rozważ odszukanie dodatkowych dokumentów szkoleniowych, które poruszają ten sam temat w tej samej kategorii i prześlij je do tej samej lub nowej kategorii.

Jeżeli wartości precyzji i poprawności są zadowalające, a dokumenty pozostają przypisane po wielu iteracjach, usuń je.

Proces wartościowania należy uruchomić ponownie, gdy zmieniono systematykę, np. usunięto kategorię, połączono dwie albo przeniesiono wiele dokumentów. Niewielkie zmiany na poziomie kategorii, np. dodawanie nowego dokumentu szkoleniowego lub kopiowanie dokumentów, mogą być wykonywane dla wielu kategorii bez konieczności ponownego wartościowania. Należy zauważyć, że nie jest dostępna funkcja służąca do wycofania wprowadzonych zmian.

Zatrzymaj wartościowanie, gdy nie ma dokumentów oznaczonych na czerwono lub niebiesko lub gdy osiągnięto precyzję i poprawność większą od 90%.

## Szkolenie katalogu

Po wykonaniu wartościowania systematyki i otrzymaniu zadowalających wyników, należy przeszkolić systematykę przy użyciu wszystkich dokumentów szkoleniowych w celu utworzenia specjalnego typu metadanych, modelu klasyfikacji, który może być kolejno używany do klasyfikowania nowych dokumentów przez usługę klasyfikacji eksploracji informacji.

Aby rozpocząć fazę szkolenia, wybierz katalog, który wymaga szkolenia, kliknij kartę **Szkolenie**, a następnie kliknij przycisk **Rozpocznij szkolenie**.

W górnej części prawej ramki wyświetlana jest nazwa wybranego katalogu oraz status jego szkolenia. Przykładowe statusy szkolenia to:

- Dokumenty nie mogą być sklasyfikowane. Aby zaktualizować klasyfikator, należy przeszkolić katalog. Jest to ustawienie domyślne dla nowo utworzonego katalogu. W takim przypadku pole **Data ostatniego szkolenia** jest puste. Albo katalog jest nowy i nigdy nie był szkolony, albo wyniki najnowszego szkolenia nie są poprawne, ponieważ np. zmieniono nazwę kategorii lub ją usunięto. Próba użycia usługi klasyfikowania do nowych dokumentów zakończy się błędem.
- Zmieniono zbiór dokumentów szkoleniowych w katalogu. Dokumenty są wciąż poklasyfikowane zgodnie z ostatnimi wynikami szkolenia. Przeszkol katalog, aby zaktualizować klasyfikowanie.
- W tym katalogu trwa szkolenie.
- Katalog jest aktualny, szkolenie nie jest wymagane.

Aby zatrzymać proces szkolenia, kliknij przycisk **Zatrzymaj szkolenie**. Jeżeli szkolenie zostanie zatrzymane, proces ten trzeba będzie zacząć od nowa.

Podczas trwania procesu szkolenia nie można dodawać do katalogu nowych dokumentów.

---

## Strojenie wydajności

Gdy w katalogu tworzone są rekordy, indeksy tekstowe atrybutów dopuszczalnych w wyszukiwaniu tekstowym, np. IKF\_CONTENT, bardzo się rozrastają, co powoduje znaczące obniżenie wydajności wyszukiwania. Aby zoptymalizować miejsce na dysku i zwiększyć wydajność, należy okresowo reorganizować indeksy tekstowe, szczególnie po wykonaniu ich znaczącej aktualizacji.

Reorganizacja indeksu tekstowego powinna być uruchamiana poza szczytowym wykorzystywaniem systemu, np. w nocy. Aby rozpocząć organizowanie indeksu, uruchom:

- w systemie Windows: `...\ikf\lcfReorg.cmd`
- w systemie AIX: `.../ikf/bin/lcfReorg.sh`
- w systemie Solaris: `.../ikf/bin/lcfReorg`

Parametry są następujące: `lcfReorg <ID_użytkownika><hasło><nazwa_bazy_danych>`

---

## Używanie programu IBM Web Crawler

W tej sekcji opisano i wyjaśniono sposób konfigurowania programu IBM Web Crawler. Program ten jest instalowany przez program instalacyjny EIP, jeżeli zaznaczone zostało pole wyboru funkcji.

Program EIP wersja 8.2 zawiera Nawigator sieci WWW, Nawigator Lotus Notes, mechanizmy pobierania danych ze zgromadzonych plików, dokumentację w formacie HTML, przykłady konfiguracji i pomocnicze programy narzędziowe. Program IBM Web Crawler (nazywany również GCS) wymaga obsługi języka Java 1.3 lub nowszego.

Program IBM Web Crawler jest napisany w języku Java i służy do nawigowania i eksplorowania informacji. Gdy zostanie on skierowany w dowolne miejsce, pobiera całą zawartość i przeszukuje ją.

Program IBM Web Crawler może pobierać zawartość z intranetu, ekstranetu, internatu, baz danych Lotus Notes (bezpośrednio albo za pomocą Domino), lokalnego systemu plików. Program IBM Web Crawler jest zaprojektowany w taki sposób, aby dodawanie nowych protokołów było proste. Dane mogą być dowolnego typu, np. HTML, załączniki Notes i pliki multimedialne.

Program IBM Web Crawler może eksplorować metadane i tekst z wielu różnych rodzajów dokumentów. Na przykład z dokumentów HTML można uzyskać:

- adres URL
- tytuł
- treść
- godzinę ostatniej modyfikacji
- metaznaczniki, np. autor, słowa kluczowe, opis itp.

Użytkownicy mogą wybierać opcje wyszukiwania ze zbioru opcji dostępnych dla określonego typu danych. Zawartość lub przeszukiwane metadane są zapisywane na lokalnym dysku. Program IBM Web Crawler może używać technologii Network Solutions Outside In do pobierania tekstu z ponad 200 różnych rodzajów danych. Program IBM Web Crawler jest również zaprojektowany w taki sposób, aby dodawanie nowych programów eksplorujących było proste.

Program IBM Web Crawler jest dostępny dla systemów Windows NT 4.0 i Windows 2000. Instalacja i konfiguracja programu IBM Web Crawler zajmuje ok. 30 minut. Na komputerze PC z procesorem 500 MHz pobieranie i eksplorowanie danych przebiega z przybliżoną prędkością 10 plików na sekundę. Program przetestowano przy użyciu 1 miliona obiektów (200000 notatek). Obsługuje on wielu użytkowników oraz wiele różnych konfiguracji dla każdego użytkownika, możliwe jest też używanie preferowanego języka.

## Możliwości programu IBM Web Crawler

Program instalacyjny instaluje następujące katalogi:

`x:/<katalog_instalacyjny>/run`

Pliki wsadowe programu IBM Web Crawler dla WWW i przykładowe konfiguracje.

`x:/<katalog_instalacyjny>/notes-run`

Pliki wsadowe programu Web Crawler dla Notes i przykładowe konfiguracje.

`x:/<katalog_instalacyjny>/lib`

Pliki .jar i .zip programu IBM Web Crawler oraz pliki filtrujące.

## Konfigurowanie i uruchamianie programu IBM Web Crawler dla sieci WWW

W tej sekcji opisano sposób konfigurowania i uruchamiania programu IBM Web Crawler dla sieci WWW. Program IBM Web Crawler używa protokołów HTTP, FTP, NNTP lub serwerów plików oraz tworzy streszczenia plików HTML i innych obiektów. Streszczenia są to pliki, jeden dla każdego dokumentu lub obiektu, zawierające metadane i pełny tekst.

### Konfiguracja podstawowa

W tej sekcji zawarto instrukcje objaśniające edytowanie plików konfiguracyjnych programu IBM Web Crawler, które zapisano w formacie XML. Aby ułatwić rozpoczęcie korzystania z programu, dołączono następujące dwa przykładowe pliki konfiguracyjne:

- Plik `config-db2.xml` służący do używania programu IBM Web Crawler z bazą danych DB2 UDB.
- Plik `config-sample.xml` służący do używania programu IBM Web Crawler bez bazy danych DB2 UDB.

1. Otwórz okno wiersza komend.
2. Przejdź do podkatalogu `run`, w którym zainstalowano program IBM Web Crawler. Na przykład, jeżeli program IBM Web Crawler zainstalowano w systemie Windows, wpisz `cd x:<cmdbroot>\gcs\run`. Jeżeli program IBM Web Crawler zainstalowano w systemie AIX, wpisz `cd /usr/lpp/cmb/gcs`.



**Wskazówka:** Należy pamiętać o zachowaniu kopii oryginalnego pliku. Błąd w pliku konfiguracyjnym może uniemożliwić używanie programu IBM Web Crawler. Podczas edytowania należy zachować ostrożność.

3. Aby uruchomić program IBM Web Crawler z bazą danych DB2 (większa skalowalność, ale nie najwyższa szybkość), zmodyfikuj plik `config-db2.xml`. Na przykład wpisz w wierszu komend `edit config-db2.xml`.
4. Aby uruchomić program IBM Web Crawler bez bazy danych DB2 (mniejsza skalowalność, ale większa szybkość), zmodyfikuj plik `config-sample.xml`. Na przykład wpisz w wierszu komend `config-sample.xml`.

Aby uruchomić odwiedzanie  $n$  adresów URL bez bazy danych, potrzeba około  $n/1000$  MB pamięci RAM na komputerze do przechowywania metadanych. Na przykład, aby odwiedzić 500000 adresów URL, potrzeba 512 MB pamięci RAM. Aby wykorzystać zalety dużej pamięci RAM, zmodyfikuj plik `crawlweb.bat` i zwiększ wartość parametru `JVMXmx`.

## Konfigurowanie Nawigatora sieci WWW dla eksploracji informacji

Aby można było używać Nawigatora sieci WWW razem z eksploracją informacji, czyli mieć możliwość zastosowania funkcji eksploracji informacji do nawigowanych dokumentów, wymagane jest podanie w plikach konfiguracyjnych przedstawionych poniżej ustawień konfiguracyjnych (`config-sample.xml` i `config-db2.xml`):

```
<globals ...
 max-urls="5000"
 temp-filepool-class="FullPathFilePool"
 summaries-dir="webspaces/ikf/disks/1/"
 summaries-filepool-class="DistributedEIPFilePool"
... >
...
</globals>
<group-list>
 <group ...>
 ...
 <summarizer-config>
 ...
 <resource-handler content-type="*"
 summarizable="EipHtmlSummarizable"
 summary-maker="EipHtmlRawSummaryMaker" />
 ...
 </summarizer-config>
 ...
 </group>
</group-list>
```

Ustawienia te są częścią przykładowego pliku konfiguracyjnego dla składnika eksploracji informacji, `im-crawler-config-sample.xml`, który jest umieszczony w następującym katalogu:

Dla systemu Windows:

```
<CMBROOT>\samples\java\beans\infomining\webcrawler\
```

Dla systemu UNIX (AIX i Solaris):

```
<CMBROOT>/samples/java/beans/infomining/webcrawler/
```

## Konfigurowanie opcji DB2 programu IBM Web Crawler

Aby skonfigurować opcję DB2, należy utworzyć bazę danych. Wymaga to posiadania uprawnień administratora DB2. Jeżeli jest to konieczne, należy zalogować się na konto administratora DB2. Nazwa bazy danych może być dowolna (ale zgodna z zasadami DB2), ale jeśli jest ona inna niż `gcs`, należy zaktualizować parametr `dbname` w pliku konfiguracyjnym.

Posiadając uprawnienia administratora bazy danych, należy uruchomić komendę w wierszu komend DB2:

```
db -createdb <uzytkownik><haslo>[baza_danych]
```

Jeżeli nie zostanie podana nazwa bazy danych, zostanie przyjęta domyślna nazwa **gcs**. Po utworzeniu bazy danych należy utworzyć w niej tabele używane przez program IBM Web Crawler:

```
db -createtables<uzytkownik><haslo>[baza_danych]
```

Programu IBM Web Crawler można używać po zakończeniu procesu tworzenia bazy danych i tabel.

Aby możliwe było używanie nowej bazy danych, w pliku konfiguracyjnym (w sekcji **urlpool-config**) wymagane są następujące ustawienia: **dbname**:

- Nazwa bazy danych (utworzonej w poprzednim kroku), np. **gcs**.
- User name: Nazwa użytkownika, np. **db2admin**.
- Password: Hasło użytkownika, np. **db2admin**.

Podane wyżej parametry muszą być właściwie ustawione. Nie należy zmieniać wielkości pamięci podręcznej ani sterownika. Edytowanie należy kontynuować w celu określenia zasięgu odwiedzania dla systemu.

## Określanie zasięgu odwiedzania

Niżej wymienione ustawienia w pliku konfiguracyjnym są wymagane do określenia zasięgu odwiedzania niezależnie od faktu używania (lub nie używania) bazy danych DB2.

Ustawienia te umieszczone są w sekcji **crawler-config**.

### **seed list**

Jeden lub kilka adresów początkowych. Adres URL musi być dostępny. Należy go zweryfikować za pomocą przeglądarki, np. <http://www.<moj-komputer>.com.pl/>

### **content-type-pattern-list**

Odwiedzane są adresy URL, umieszczone na stronie, których rozszerzenie pasuje do podanego wzorca, np. **htm\***.

### **include-pattern-list**

Odwiedzane są adresy URL, umieszczone na stronie, które pasują do podanego wzorca, np. **<wroc.eu.pl>**.

Można również ustawić następujące pozycje:

### **recursion-depth**

Maksymalna odległość w odsyłaczach, które mają być odwiedzane od dowolnego punktu początkowego. Aby określić nieskończoną odległość, należy podać **-1**.

### **exclude-pattern-list**

Określa, że odwiedzane będą adresy URL, które nie pasują do podanego wzorca, np. **\*cgi-bin\***.

### **Właściwości systemowe**

Jeżeli odwiedzanie inicjowane jest z komputera umieszczonego za firewallem (zaporą sieciową), na którym nie ma możliwości skorzystania z mechanizmu SOCKS, to należy ustawić wartość **socksProxy**.

## Uruchamianie programu IBM Web Crawler

Po zakończeniu edytowania pliku **.xml** należy zapisać go.



Aby uruchomić program IBM Web Crawler, użyj programu wsadowego `crawlweb` i pliku konfiguracyjnego. Otwórz okno wiersza komend i wpisz:

- w systemie Windows: `crawlweb.bat <PLIK_KONFIGURACYJNY>`
- w systemie AIX: `crawlweb.sh <PLIK_KONFIGURACYJNY>`

Aby uruchomić program z bazą danych DB2, wpisz: `crawlweb config-db2.xml` i naciśnij Enter. Aby uruchomić bez bazy danych DB2, wpisz: `crawlweb config-sample.xml` i naciśnij Enter.

**Wskazówka:** Zaplanuj regularne raportowanie odwiedzania. Po odwiedzeniu streszczenia są zapisywane w katalogu określonym przez `summaries-dir`. Domyślnie streszczenie zapisuje oryginalny obiekt wraz z metadanymi jako pliki `.html` w drzewie. Plik protokołu można odczytywać podczas odwiedzania albo później, aby uzyskać dodatkowe informacje.

### Konfiguracja zaawansowana

W tej sekcji omówione zostaną opcje konfiguracyjne. Przykład konfiguracji zapisany jest w pliku `config-sample2.xml` w Rozdział 7, “Przykładowe pliki programu IBM Web Crawler”, na stronie 105. Przykład ilustruje konfigurację następujących składników:

- nawigator i wątki streszczające;
- monitor graficzny;
- opcje protokołowania;
- SOCKS;
- odwiedzanie baz Lotus Domino;
- wiele typów danych;
- wyłączenia;
- użycie `InsoSummarizable` do otrzymania streszczeń obiektów, takich jak pliki `.pdf`.

Formalna definicja dostępnych parametrów zamieszczona jest w pliku `config.dtd`. **Zalecenie:** Nie edytuj tego pliku. Utwórz jego kopię pod inną nazwą.

## Plik konfiguracyjny programu IBM Web Crawler

Plik konfiguracyjny jest plikiem XML, który mówi programowi IBM Web Crawler, jakie zasoby WWW należy pobierać i jak je streszczać. W tej sekcji opisano każdy element i atrybut, który może być ustawiany w pliku `config.xml`. Aby uzyskać informacje na temat używania programu IBM Web Crawler z bazami Notes, patrz “IBM Web Crawler dla Notes” na stronie 89.

Program IBM Web Crawler sprawdza zgodność zawartości pliku konfiguracyjnego z plikiem `gcs-config.dtd`. Jeżeli wystąpią poważne błędy, np. nie podano żadnego adresu URL, program IBM Web Crawler wyświetla komunikat o błędzie i kończy pracę. Mniej istotne błędy, np. nieznany atrybut lub wartość, są raportowane w pliku protokołu, nie powodując przerywania pracy programu. **Zalecenie:** Przed edycją pliku konfiguracyjnego należy zachować jego oryginalną kopię. Błąd w pliku konfiguracyjnym może uniemożliwić używanie programu IBM Web Crawler.

Przykładowe pliki konfiguracyjne są dostarczane z programem IBM Web Crawler.

### Plik `<gcs-config>`

Plik `gcs-config` składa się z dwóch sekcji: **globals** i **group-list**. Rozdział 7, “Przykładowe pliki programu IBM Web Crawler”, na stronie 105 zawiera przykład pliku `gcs-config`.

**globals** Element `globals` pobiera ustawienia programu IBM Web Crawler, takie jak system plików, wydajność i informacje o sieci.

**group-list**

Element group-list konfiguruje odwiedzanie i streszczanie grup, gdzie grupą jest zbiór zasobów, takich jak domena.

**Element <globals>**

Element globals reprezentuje globalne ustawienia programu IBM Web Crawler. Ustawienia są kodowane jako atrybuty globalne i elementy potomne.

Następująca lista definiuje atrybuty globalne. "Parametr <logger-config>" na stronie 79 zawiera definicje elementów potomnych.

**max-urls**

Maksymalna liczba odwiedzanych adresów URL. Jest to dodatnia liczba całkowita, wartość domyślna: 100000.

**summaries-dir**

Katalog, w którym zapisywane są streszczenia zasobów. Wartość domyślna: summaries/.

**summaries-filepool-class**

Typ puli plików używanych przez streszczenia zasobów. Określa sposób nazywania plików streszczeń oraz strukturę (jeżeli istnieje) podkatalogów. Wartość domyślna: FullPathFilePool, która tworzy podkatalogi dla hosta, a następnie używa takiej samej struktury podkatalogów i nazw plików jak w adresie URL.

**num-crawlers**

Liczba używanych wątków pobierania. Jest to dodatnia liczba całkowita, wartość domyślna: 20.

**num-summarizers**

Liczba używanych wątków streszczających. Jest to dodatnia liczba całkowita, wartość domyślna: 5. Zalecenia dotyczące ustawienia parametrów num-crawlers i num-summarizers:

1. Ustaw liczbę wątków pobierania na wartość obliczoną w następujący sposób:  $\text{szybkość\_komputera\_MHz}/20$ . Na przykład dla komputera z procesorem 600 MHz, wpisz 30.
2. Ustaw liczbę wątków streszczających na wartość obliczoną w następujący sposób:  $\text{szybkość\_komputera\_MHz}/80$ . Na przykład dla komputera z procesorem 600 MHz, wpisz 8.
3. Uruchom proces, obserwując panel wydajności Menedżera zadań. Jeżeli użycie procesora *kiedykolwiek* przez dłużej niż 1s będzie na poziomie 100%, powróć do kroku nr 1 i ustaw mniejszą liczbę, np. 3/4 poprzedniej wartości.

Jeżeli podczas prób zaobserwujesz w monitorze tekstowym, że dla wątków streszczających liczby zadań do wykonania są znacznie niższe od liczby skonfigurowanych wątków streszczających, możesz zmniejszyć ich liczbę (im mniej, tym lepiej) i jednocześnie zwiększyć liczbę wątków pobierania (im więcej, tym lepiej), uważając jednocześnie, aby nie przekroczyć warunku podanego w kroku 3. Aby uzyskać najlepsze wyniki, należy zastosować najszybsze dostępne połączenia sieciowe i umieścić podsumowania, bazę danych, obszar tymczasowy i protokoły na osobnych dyskach.

**text-monitor**

Jeżeli parametr ten jest ustawiony, monitor tekstowy wyświetla status programu IBM Web Crawler co pięć sekund na standardowym wyjściu. Jeżeli parametrowi nadana jest wartość dziesiętna, to jest ona używana jako liczba sekund między kolejnymi odświeżeniami. Domyślnie parametr ten nie jest ustawiony.

**graph-monitor**

Jeżeli parametr ten jest ustawiony, program IBM Web Crawler wyświetla swój status

w postaci graficznej. Jeżeli parametrowi nadana jest wartość dziesiętna, to jest ona używana jako liczba sekund między kolejnymi odświeżeniami. Domyślnie parametr ten nie jest ustawiony.

**log-file** Określa główny plik protokołu. Wartość domyślna: log/log.txt

**Wskazówka:** Dodatkowy program protokołu można określić w elemencie logger-config.

**log-priority**

Ustawia priorytet protokołowania. Możliwe wartości: info, warn (ostrzeżenie) lub error (błąd). Wartość domyślna: warn.

**Wskazówka:** Dodatkowy program protokołu można określić w elemencie logger-config.

**temp-dir**

Katalog, w którym zapisywane są pliki tymczasowe. **Wskazówka:** Wszystkie pliki w tym katalogu mogą być usunięte przez program IBM Web Crawler. Nie należy zmieniać domyślnego ustawienia x:/temp/gcs.

**temp-filepool-class**

Typ puli plików używanej dla plików tymczasowych. **Zalecenie:** Nie należy zmieniać domyślnego ustawienia TempFilePool.

**content-dir**

Katalog, w którym program IBM Web Crawler zapisuje pliki danych. Zwykle parametr content-dir ma taką samą wartość jak parametr temp-dir.

**content-filepool-class**

Typ puli plików używanej dla plików danych. Zwykle jest to taka sama wartość jak parametru temp-filepool-class.

**how-often-to-gc**

Liczba odwiedzonych adresów URL między kolejnymi żądaniami oczyszczania pamięci. **Zalecenie:** Zaleca się zdefiniowanie wartości całkowitej  $\geq 50$ . Wartość domyślna: 100.

**max-resource-pool-size**

Maksymalna wielkość kolejki zasobów oczekujących na streszczenie. **Zalecenie:** Zaleca się zdefiniowanie wartości całkowitej  $\geq 10$ . Wartość domyślna: 10 oczekujących zasobów na jeden program streszczający.

**connect-timeout**

Definiuje liczbę milisekund oczekiwania przed uznaniem połączenia sieciowego jako nieudanego. Wartość domyślna: 4000. Poprawny zakres wartości: 1000-60000.

**read-timeout**

Definiuje liczbę milisekund oczekiwania przed uznaniem operacji odczytu jako nieudanej. Wartość domyślna: 6000. Poprawny zakres wartości: 1000-60000.

**cookies**

Definiuje, czy należy sprawdzać informacje cookie w nagłówku HTTP i zapisywać je w bazie danych. Domyślnie parametr ten nie jest ustawiony. Można go włączyć, ustawiając wartość na 'on'.

**locale** Definiuje język dla streszczania i protokołowania. Wartość domyślna: en\_US.

Globalne elementy potomne zawierają logger-configs, urlpool-config i system-properties.

## Parametr <logger-config>

Plik logger-config umożliwia zaawansowane sterowanie protokołowaniem, sposobem jego formatowania i pliku, w którym protokół jest zapisywany. Parametry default log-file i

log-priority są określane jako atrybuty globalne. “Protokołowanie w programie IBM Web Crawler” na stronie 85 zawiera więcej informacji na temat protokołowania.

**category**

Kategoria programu protokołującego, który jest konfigurowany, np. `gcs.crawler`. Jeżeli nie jest określona, konfigurowany jest domyślny program protokołujący. Należy pamiętać o tym, że ustawienia określonej kategorii mają wpływ na wszystkie podkategorie.

**priority**

Minimalny priorytet protokołowanego komunikatu. Jeżeli nie jest określony, priorytet jest określany z kategorii nadrzędnej (albo przyjmowana będzie wartość log-priority z sekcji globalnej).

**log-file**

Definiuje, gdzie zapisywać plik protokołu. Jeżeli rozpoczyna się znakiem '+', to plik ten będzie używany oprócz zdefiniowanych nadrzędnych plików protokołów (jeżeli są zdefiniowane). Jeżeli parametr nie jest podany, użyty zostanie nadrzędny plik protokołu (albo przyjmowana będzie wartość log-file z sekcji globalnej).

**Wskazówka:** Nie należy definiować tego samego pliku dla wielu programów protokołujących, ponieważ będą wzajemnie go nadpisywać.

**log-layout**

Definiuje układ używany dla każdego komunikatu zapisywanego w pliku protokołu.

**Parametr <urlpool-config>**

Plik `urlpool-config` konfiguruje składnik programu IBM Web Crawler, w którym zapisywane są adresy URL. Istnieje kilka opcji dla puli adresów URL. Można ją przechowywać w pamięci, w bazie danych DB2, można użyć specjalnej wersji zużywającej mniej pamięci, która nie przechowuje wszystkich informacji na temat adresu URL. Jeżeli parametr `urlpool-config` nie zostanie określony, pula adresów URL przechowywana jest w pamięci. Parametr `urlpool-config` może mieć elementy potomne `urlpool-param`, określające dodatkowe parametry, np. informacje dotyczące bazy danych.

**urlcontainer-class**

Typ używanego pojemnika adresów URL. Można podać:

- `DB2URLContainer`, aby używać bazy danych DB2;
- `MemoryURLContainer`, aby nie używać bazy danych DB2 (wartość domyślna);
- `BigMemoryURLContainer`, aby nie używać bazy danych DB2 i zapisywać w pamięci dodatkowe informacje na temat adresów URL.

**urlcollection-class**

Typ używanej kolekcji adresów URL. Można podać:

- `DB2URLCollection`, aby używać bazy danych DB2;
- `MemoryURLCollection`, aby nie używać bazy danych DB2 (wartość domyślna);
- `BigMemoryURLCollection`, aby nie używać bazy danych DB2 i zapisywać w pamięci dodatkowe informacje na temat adresów URL.

**Parametr <urlpool-param>**

Używany do przekazywania parametrów dla `urlcollection-class`. Przykład umieszczono w informacjach na temat połączenia w konfiguracji używającej bazy danych DB2, zawiera go Rozdział 7, “Przykładowe pliki programu IBM Web Crawler”, na stronie 105.

**Name** Definiuje nazwę parametru.

**value** Definiuje wartość parametru.

**Wskazówka:** Należy zachować ostrożność podczas używania tych parametrów, ponieważ nie jest sprawdzana ich poprawność.

### Parametr <system-properties>

Właściwości systemu reprezentują listę ustawień właściwości systemu.

#### <właściwość>

Przykładowa konfiguracja zaawansowana znajduje się w definicji bramy SOCKS.

**Name** Nazwa parametru.

**value** Wartość parametru.

Alternatywnie można skonfigurować program IBM Web Crawler do używania bramy PROXY do dostępu do innych serwerów w następujący sposób:

```
<system-properties>
 <property name="proxySet" value="true"/>
 <property name="proxyHost" value="proxy.nazwa_hosta"/>
 <proxy port name="proxyPort" value="80"/>
</system-properties>
```

**Wskazówka:** Nie są sprawdzane błędy w tych parametrach, należy więc zachować ostrożność.

### Parametr <group-list>

Parametr group-list definiuje listę jednej lub kilku grup elementów.

#### <group>

Element grupowy reprezentuje pojedynczą grupę zasobów, która będzie odwiedzana i streszczana w podobny sposób. Każda grupa musi mieć niepowtarzalny atrybut name i przynajmniej jeden element potomny crawler-config, określający elementy do odwiedzania. Jeżeli nie mają być używane domyślne programy streszczające, można określić potomny element summarizer-config. **Wskazówka:** Używanie nakładających się grup (takie same adresy URL w kilku grupach) może prowadzić do uzyskania nieoczekiwanych wyników. Wielogrupowe adresy URL są kojarzone tylko z pierwszą grupą, w której zostały znalezione.

**name** - niepowtarzalna nazwa grupy (wymagane).

### Parametr <crawler-config>

Do określania zasięgu odwiedzania używany jest następujący algorytm. Program pobiera każdy adres URL z listy seed-list, analizuje adresy URL określone w treści i dodaje je do listy to-be-crawled, jeżeli pasują one do:

- przynajmniej jednej reguły określonej w content-type-pattern-list
- przynajmniej jednej reguły określonej w include-pattern-list
- do żadnej z reguł określonej w exclude-pattern-list.

Parametr crawler-config wymaga określenia jednego atrybutu: **recursion-depth**. Wyznacza on liczbę poziomów dla każdego adresu podanego w parametrze seed, który można odwiedzić. Wartością domyślną jest -1, co oznacza nieskończoną głębokość odwiedzania.

### Parametr <seed-list>

Jest to lista adresów URL. Każdy z nich może zawierać informacje uwierzytelniające.

#### Parametr <seed>

Reprezentuje adres URL, od którego rozpoczyna się odwiedzanie (może zawierać informacje uwierzytelniające). Każdy adres musi być podany w pełnej postaci, np. <http://<serwer>.com.pl/>. Należy unikać adresów, do których dostęp jest ograniczony, które są niedostępne albo wskazują na strony bez tekstu. Przydatne może być ustawienie adresu

strony, w której określono listę adresów do odwiedzenia. Taka lista jest łatwa do aktualizacji, przeglądania i testowania za pomocą przeglądarki.

**URL** Adres URL, od którego rozpoczyna się odwiedzanie.

### **Parametr <authentication>**

Opcjonalne uwierzytelnianie wysyłane dla adresu URL chronionego przez *podstawowe uwierzytelnianie* zdefiniowane w RFC 2617.

#### **username**

Nazwa użytkownika.

#### **password**

Hasło użytkownika.

Na przykład:

```
<seed url="http://moj-serwer.com.pl/"><authentication username="janusz"
password="PankotekPlutoJestChory"/></seed>
```

### **Parametr <content-type-pattern-list>**

Jest to lista wzorców używanych do włączania typów danych do odwiedzania, określanych przez rozszerzenie. Dowolne rozszerzenie pliku w adresie URL (.html, .gif, .rpm itp.), które jest zgodne z url-name-pattern na tej liście, przechodzi ten test. Adresy URL, które nie mają rozszerzenia, przechodzą test. Jeżeli parametr content-type-pattern-list nie jest określony albo jest pusty, to akceptowane są adresy URL bez rozszerzeń.

### **Parametr <include-pattern list>**

Jest to lista wzorców służących do określania, które adresy URL będą odwiedzane. Każdy adres URL, który jest zgodny z dowolnym parametrem url-obj-pattern, url-regex-pattern, url-name-pattern lub url-predicate-pattern, przejdzie test pomyślnie. Jeżeli parametr include-pattern-list nie jest określony albo jest pusty, wtedy wszystkie adresy URL będą przyjmowane.

### **Parametr <exclude-pattern list>**

Jest to lista wzorców służących do określenia, które adresy URL nie będą odwiedzane. Każdy adres URL, który jest zgodny z dowolnym parametrem url-obj-pattern, url-regex-pattern, url-name-pattern lub url-predicate-pattern, nie będzie odwiedzany. Jeżeli parametr exclude-pattern-list nie jest określony albo jest pusty, żaden adres URL nie będzie odrzucony.

### **Parametr <url-obj-pattern>**

Jest to wzorec, który może być używany do porównywania z różnymi częściami adresu URL (protokół, nazwa serwera, nazwa pliku) z możliwością używania znaków zastępczych. Może być używany z parametrami exclude-pattern-list i include-pattern-list. Wzorec w każdej części może mieć na początku lub końcu znak zastępczy '\*', który oznacza dowolny ciąg znaków. Nie można używać znaków zastępczych w środku wzorca. Wielkość znaków nie jest rozróżniana. Każda pominięta część adresu URL oznacza automatycznie zgodność.

Poniżej omówiono sposób podziału przykładowego adresu URL

[http://www.ibm.com/products/index.html?query#ref:](http://www.ibm.com/products/index.html?query#ref)

- Protokół: http
- Nazwa hosta: www.ibm.com
- Port: -1 (nie został określony)
- Pełna nazwa pliku: /products/index.html?query
- Ścieżka: /products/index.html
- Katalog: /products/



- Nazwa pliku: index.html
- Rozszerzenie: .html
- Zapytanie: query
- Referencja: ref

Poniżej omówiono dokładniej każdy z elementów parametru url-obj-pattern.

#### protokół

Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z protokołem podanym w adresie URL, np. http.

**host** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z nazwą hosta podaną w adresie URL, np. \*.ibm.com.

**port** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z numerem portu podanym w adresie URL, np. 80.

**file** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z nazwą pliku podaną w adresie URL, np. \*.htm\*. Nazwa pliku rozpoczyna się pierwszym znakiem ukośnika po nazwie hosta i może zawierać zapytanie, ale nie może zawierać referencji. *Pełną nazwą pliku* w adresie URL http://www.ibm.com/products/index.html?query#ref jest /products/index.html?query.

**path** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny ze ścieżką podaną w adresie URL, np. \*.html. Ścieżka rozpoczyna się pierwszym znakiem ukośnika po nazwie hosta i nie może zawierać zapytania ani referencji. *Ścieżką* w adresie URL http://www.ibm.com/products/index.html?query#ref jest /products/index.html.

**dir** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z katalogiem podanym w adresie URL, np. /produkty/. Katalog jest częścią ścieżki, rozpoczyna się od pierwszego znaku ukośnika i kończy na ostatnim znaku ukośnika. *Katalogiem* w adresie URL http://www.ibm.com/products/index.html?query#ref jest /products/. Nie zawiera on zapytania, ani referencji. Należy zauważyć, że źle zbudowane adresy URL, które nie zawierają ostatniego znaku ukośnika, np. http://www.ibm.com/products, nie będą pasowały do katalogu. W przykładowym źle zbudowanym adresie URL katalogiem jest /.

#### filename

Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z nazwą pliku, np. index.html. Nazwa pliku jest częścią ścieżki występującą po ostatnim znaku ukośnika. *Nazwą pliku* w adresie URL http://www.ibm.com/products/index.html?query#ref jest index.html. Nie zawiera on zapytania, ani referencji.

#### extension

Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z rozszerzeniem nazwy pliku podanym w adresie URL, np. htm\*. Zalecane jest używanie content-type-pattern-list, jeśli jest to możliwe.

**query** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z zapytaniem podanym w adresie URL.

**ref** Wzorzec (może zawierać znaki zastępcze), który musi być zgodny z referencją podaną w adresie URL (nie jest używane w HTTP). Na przykład <url-obj-pattern host="\*.ibm.com"/> będzie zgodne z każdą stroną HTML firmy IBM.

### Parametr <url-regex-pattern>

Parametr url-regex-pattern jest to wzorzec, który jest porównywany z adresem URL z wykorzystaniem wyrażeń regularnych. Może on być używany zarówno w

exclude-pattern-list jak i w include-pattern-list. Używa on pakietu com.ibm.regex (regex4j) i obsługuje większość funkcjonalności wyrażeń regularnych używanych w języku Perl 5. Można określić dwa wyrażenia regularne, jedno, które *musi* pasować i drugie, które *nie może* pasować. Można podawać inne opcje, takie jak i, oznaczające brak rozróżniania wielkości liter. Więcej informacji zawiera publikacja [Regex4j Regular Expressions](#).

**match** Wyrażenie Perl, z którym adres URL musi być zgodny.

**no-match**

Wyrażenie Perl 5, z którym adres URL nie może być zgodny.

**options**

Opcjonalne modyfikatory, np. i oznacza brak rozróżniania wielkości liter.

Na przykład wyrażenie `<url-regex-pattern match="^http://www\.ibm\.com/.*\.html?$"/>` będzie pasować do wszystkich stron HTML na głównym serwerze WWW firmy IBM.

**Parametr <url-name-pattern>**

To jest prosty wzorzec ze znakami zastępczymi, używany do porównywania pełnych adresów URL lub rozszerzeń plików. Może być używany w `content-type-pattern-list` i `include-pattern-list` oraz w `exclude-pattern-list`. Znakiem zastępczym jest znak '\*', który oznacza dowolny ciąg znaków. Znak zastępczy musi być umieszczony na początku lub końcu wzorca. Nie może występować w środku wzorca. Wielkość znaków nie jest rozróżniana.

Na przykład `<url-name-pattern name="*.ibm.com/*"/>` będzie zgodne ze wszystkimi plikami na stronie firmy IBM, ale `<url-name-pattern name="*.ibm.com/*.html"/>` jest niepoprawne, ponieważ znak zastępczy występuje w środku wzorca.

**name** Wzorzec (może opcjonalnie zawierać znak zastępczy '\*' na początku lub końcu), który musi być zgodny z łańcuchem adresu URL.

**Parametr <url-predicate-pattern>**

Ten wzorzec powoduje załadowanie klasy Java o nazwie UnaryPredicate, która jest używana do porównywania zgodności adresów URL. Może być używany z parametrami `exclude-pattern-list` i `include-pattern-list`. Klasa musi mieć zdefiniowaną metodę `public boolean execute(URL url)`, która zwraca prawdę, jeżeli URL jest zgodny z predykatem.

**class** W pełni określona nazwa klasy UnaryPredicate.

**Parametr <summarizer-config>**

Jest to konfiguracja programu streszczającego z listą potomnych elementów `resource-handler`. W grupie może wystąpić tylko jeden element `summarizer-config`.

**<resource-handler>**

Określa typ streszczenia dla zasobu (np. strony WWW lub artykułu z grupy dyskusyjnej) w oparciu o typ zawartości, np. (text/html) lub rozszerzenie nazwy pliku (htm). Gdy zasób jest gotowy do streszczenia, program IBM Web Crawler sprawdza elementy `resource-handler` w kolejności wystąpienia i używa pierwszego, który pasuje do typu zawartości lub rozszerzenia pliku. Jeżeli żaden nie będzie pasować, domyślnie użyte zostaną `Copy2RdfSummarizable` i `Copy2RdfSummaryMaker`. Można zmienić to zachowanie, definiując na końcu listy `resource-handler` bez `content-type` lub `file-extension`.

Parametr `resource-handler` może posiadać elementy potomne `summarizer-param`, które służą do przekazywania specjalnych parametrów do klasy `SummaryMaker`.

**content-type**

Wzorzec (może zawierać znaki zastępcze), który musi pasować do `content-type` zasobu, np. `*htm*`.



**file-extension**

Wzorzec (może zawierać znaki zastępcze), który musi pasować do rozszerzenia pliku zasobu, np. htm\*.

**summarizable**

Nazwa klasy zasobu Summarizable, np. HtmlRawSummarizable.

**summary-maker**

Nazwa klasy SummaryMaker, np. HtmlRawSummaryMaker.

We wzorcach content-type i file-extension dozwolone jest użycie znaków zastępczych. Znakiem zastępczym jest znak '\*', który oznacza dowolny ciąg znaków. Znak zastępczy musi być umieszczony na początku lub końcu wzorca. Nie można używać znaków zastępczych w środku wzorca. Wielkość znaków nie jest rozróżniana.

Parametr resource-handler będzie zgodny, jeżeli zarówno content-type jak i file-extension będą zgodne. Nieokreślony wzorzec zawsze jest zgodny. Np.

```
<resource-handler content-type="*htm*"
summarizable="*HtmlRawSummarizable" summary-maker="HtmlRawSummaryMaker"/>
```

będzie zgodne ze wszystkimi plikami typu text/html, niezależnie od rozszerzenia nazwy pliku.

W parametrach summarizable i summary-maker nie ma potrzeby określania pełnej ścieżki klasy, jeżeli znajduje się ona w pakiecie com.ibm.IBM Web Crawler.summarizer.resource.

**Parametr <summarizer-param>**

Te parametry specjalne są przekazywane do klasy SummaryMaker. Składnia zależna jest od tej klasy.

**Name** Nazwa parametru.

**value** Wartość parametru.

**Wskazówka:** Nie są sprawdzane błędy w tych parametrach, należy więc zachować ostrożność.

## Protokołowanie w programie IBM Web Crawler

Niniejsza sekcja jest wprowadzeniem do funkcji protokołowania programu IBM Web Crawler.

Zapewnia IBM Web Crawler pełną kontrolę nad tym, co powinno być protokołowane, gdzie powinno być protokołowane i w jaki sposób powinno być sformatowane. Na przykład można wybrać zapisywanie kodów odpowiedzi każdej odwiedzonej strony do jednego pliku, status programu IBM Web Crawler (liczba odwiedzonych adresów URL, liczba działających wątków itp.) do innego pliku, streszczone adresy URL do trzeciego pliku, wszystkie ostrzeżenia programu IBM Web Crawler do czwartego pliku, a wszystkie komunikaty narzędzi z pakietu sieciowego do jeszcze innego pliku.

“Analiza przykładowego pliku protokołu programu IBM Web Crawler” na stronie 107 zawiera przykład pliku z analizą.

**Zastosowanie protokołów**

Protokołowanie jest użyteczne w rozliczaniu sieci/WWW/odwiedzania/streszczania, komunikowaniu z innymi składnikami aplikacji i debugowaniu programu IBM Web Crawler.

Rozliczanie odwiedzania i eksploracji może pomóc w uzyskaniu wielu różnych ciekawych informacji, takich jak źle skonfigurowane serwery, brakujące strony, liczba obiektów każdego typu. Skrypt `loganalysis.pl` napisany w języku Perl zawiera przykład rozliczania pliku protokołu. Aplikacje mogą potrzebować informacji z programu IBM Web Crawler, np. kiedy dane zostały usunięte.

## Konfigurowanie programów protokołujących

W pliku konfiguracyjnym programu IBM Web Crawler można określić konfigurację jednego lub kilku programów protokołujących. Atrybuty `log-priority` i `log-file` elementu `globals` określają domyślną strategię protokołowania.

Aby rozszerzyć strategię protokołowania, należy utworzyć wyrażenia `logger-config`, jako elementy potomne elementu `globals`. Każdy element określa podzbiór komunikatów programu IBM Web Crawler, kieruje je do określonego pliku i zapisuje, wykorzystując określony format. Podzbiór komunikatów jest określany na podstawie priorytetu i atrybutów kategorii. Możliwe wartości priorytetów: `trace`, `debug`, `info` i `warn` (wielkość znaków nie jest rozróżniana).

- Ustawienie wartości priorytetu określa poziom szczegółowości protokołu (`trace` jest najbardziej szczegółowym poziomem).
- `trace` i `debug` są poziomami serwisowymi - komunikaty są wyłącznie dostępne w języku angielskim.
- `info` i `warn` są poziomami dostępnymi dla użytkowników - obsługiwane są różne wersje językowe.
- Priorytet `info` tworzy wiele komunikatów. Liczbę komunikatów można zmniejszyć, określając priorytet `warn`.

## Przykłady konfiguracji protokołowania

**Protokołowanie odsyłaczy od/do, bez informacji o dacie/godzinie/wątku do pliku `log/fromto.txt`**

```
<logger-config category="gcs.url.fromto" priority="info"
log-layout="%m\n" log-file="log/fromto.txt"/>
```

**Protokołowanie streszczonych obiektów do pliku `log/resources.txt`**

```
logger-config category="gcs.summaries.list.resource"
priority="info" log-file="log/resources.txt"/
```

**Protokołowanie pominiętych adresów URL wraz z przyczyną pominięcia**

```
<logger-config category="gcs.url.skipped"
priority="info" log-file="log/urls_skipped.txt"/>
```

**Protokołowanie specjalnie przetworzonych kodów odpowiedzi HTTP**

```
<logger-config category="gcs.http.302"
priority="info" log-file="log/urls_redirected.txt"/>
<logger-config category="gcs.http.404"
priority="info" log-file="log/urls_not_found.txt"/>
```

**Protokołowanie wszystkich komunikatów, łącznie z ich priorytetem, w kategorii streszczenia**

```
<logger-config category="gcs.summarizer"
priority="TRACE" log-file="summarizer_trace.txt"
log-layout="%d: %t: %c: %p: %m\n"/>
```

## Rozwiązywanie problemów

Po napotkaniu problemu pierwszą czynnością, jaką należy wykonać, jest sprawdzenie:

**Czy strony zdefiniowane na liście seed są osiągalne?**

Strona musi istnieć (należy unikać stron przekierowujących) i musi być dostępna z systemu (bezpośrednio lub przez serwer SOCKS).

**Czy strony zdefiniowane na liście są zwykłymi stronami HTML?**

Ramki, strony WWW używające elementów flash, javascript i innych elementów nie są dobrymi kandydatami do umieszczenia na liście. Należy wybierać zwykłe strony HTML.

**Jeżeli używana jest baza danych DB2, czy było już wykonywane odwiedzanie?**

W bazie danych DB2 zapisywane są informacje o odwiedzanych stronach. Jeżeli przetworzone zostały wszystkie adresy, baza danych DB2 nie dostarczy nowych adresów. Aby rozpocząć nowe wyszukiwanie, należy użyć komendy `db -emptytables`.

**Jeżeli używana jest baza danych DB2, czy pozycje w pliku konfiguracyjnym określające sposób dostępu do bazy danych, są poprawne?**

Jeżeli połączenie z bazą danych zostanie przerwane, odwiedzanie zakończy się błędem.

**Czy plik konfiguracyjny został dokładnie sprawdzony?**

Błędy mogą spowodować uniemożliwienie używania programu IBM Web Crawler. Czy sprawdzone zostały restrykcyjne wartości parametrów `max-urls` i `recursion-depth`?

**Problem nie został usunięty?**

Należy zmodyfikować plik konfiguracyjny i zmienić priorytet protokołowania na "debug". Następnie należy uruchomić program Web Crawler i po jego zatrzymaniu sprawdzić zawartość pliku protokołu.

## Wybór programów streszczających

Przeznaczeniem programu streszczającego jest pobranie zasobu (np. strony WWW) lub hosta (np. serwera WWW) i utworzenie pliku, zawierającego informacje interesujące dla użytkownika w formacie, który jest łatwy do użycia.

Program IBM Web Crawler zawiera kilka programów streszczających, które obsługują różne typy danych, pobierają różnego rodzaju dane z zasobu i tworzą dane wyjściowe w różnych formatach. W tej sekcji opisano funkcje i wymagania dostępnych programów streszczających. Jeżeli żadna z tych funkcji nie jest wystarczająca, można napisać własny program streszczający.

W programie IBM Web Crawler dostępne są dwa rodzaje programów streszczających. Pierwszy to program streszczający pojedyncze zasoby, np. strony WWW, a drugi to program streszczający hosta, np. serwer WWW. W obecnej wersji programu można konfigurować jedynie programy streszczające zasoby.

### Wybieranie programu streszczającego zasoby

Podczas wybierania lub tworzenia programu streszczającego zasoby należy rozważyć:

- Jaki jest format pliku wejściowego? Strona w języku HTML, plik PDF, dokument WordPro lub plik XML.
- Jakie metadane są potrzebne? Nagłówek HTTP, tytuł, odsyłacze z opisem, tekst treści.
- Jaki format wyjściowy jest potrzebny? XML, HTML, RDF.

Wyboru programu streszczającego dla określonego typu zasobu dokonuje się za pomocą elementu `resource-handler` w pliku konfiguracyjnym programu IBM Web Crawler. Najpierw należy określić typ zawartości lub rozszerzenie nazwy pliku, dla którego będzie używany program streszczający. Następnie należy określić klasy Java `Summarizable` i `SummaryMaker`,

które wykonają całą pracę. Klasa summarizable reprezentuje zasób, który ma być streszczany, a klasa summary-maker reprezentuje typ streszczenia, które ma być wykonane.

### **Domyślny program streszczający (kopiowanie + RDF)**

Jest to program streszczający, który jest stosowany do każdego obiektu, którego typ zawartości nie został bezpośrednio skonfigurowany do przetwarzania przez inny program streszczający. Ten program streszczający może być używany z zasobem dowolnego typu i zapisuje dwa pliki. Pierwszy plik jest dokładną kopią oryginalnego zasobu, a drugi jest streszczeniem RDF, zawierającym oryginalny adres URL, nazwę pliku i informacje nagłówka HTTP. Program ten może być bezpośrednio skonfigurowany za pomocą parametrów DefaultSummarizable i Copy2RdfSummaryMaker.

### **Programy streszczające dla stron HTML (program streszczający surowy HTML)**

Dla zasobów HTML program ten tworzy kopię pliku z adresem URL i nagłówkiem HTTP umieszczonym na początku pliku w komentarzu. Program ten można skonfigurować za pomocą parametrów HtmlSummarizable i HtmlRawSummaryMaker.

```
<resource-handler content-type="*htm*"
 summarizable="HtmlSummarizable"
 summary-maker="HtmlRawSummaryMaker" />
```

### **Programy streszczające dla stron HTML (program EIP streszczający HTML)**

Dla zasobów HTML program ten tworzy kopię pliku z adresem URL i nagłówkiem HTTP umieszczonym na początku pliku w komentarzu. Jest on konfigurowany za pomocą parametrów EIPHtmlSummarizable i EIPHtmlRawSummaryMaker.

```
<resource-handler content-type="*htm*"
 summarizable="EIPHtmlSummarizable"
 summary-maker="EIPHtmlRawSummaryMaker" />
```

### **Programy streszczające dla stron HTML (niezapisujący program streszczający HTML)**

Ten program streszczający odwiedza strony HTML i przechodzi do umieszczonych w nich odsyłaczy, ale nie zapisuje żadnych informacji. Może on być użyteczny w sytuacji, w której np. należy streścić wszystkie pliki PDF umieszczone na serwerze (za pomocą programów streszczających INSO - XML) z pominięciem plików HTML. Program ten jest konfigurowany za pomocą parametrów HtmlSummarizable i NoWriteSummaryMaker.

```
<resource-handler content-type="*htm*"
 summarizable="InsoSummarizable"
 summary-maker="InsoSummaryMaker" />
```

### **Programy streszczające dla innych typów danych (INSO - XML)**

Ten program streszczający utworzy streszczenie XML dla jednego lub kilku z 200 rodzajów zasobów, takich jak dokumenty Microsoft Word, pliki PDF, prezentacje PowerPoint i inne. Tekst treści jest pobierany przez filtry INSO firmy Network Solutions (wymagane jest posiadanie licencji INSO). Program ten jest konfigurowany za pomocą parametrów InsoSummarizable i InsoSummaryMaker.

```
<resource-handler content-type="pdf"
 summarizable="InsoSummarizable"
 summary-maker="InsoSummaryMaker" />
```

### **Inne programy streszczające**

Jeżeli mają być streszczane inne zasoby, eksplorowane inne dane lub dane mają być tworzone w innych formatach, należy skontaktować się z firmą IBM, aby utworzyć inny program streszczający.

## IBM Web Crawler dla Notes

W tej sekcji opisano sposób konfigurowania i uruchamiania programu IBM Web Crawler dla Notes. Program IBM Web Crawler dla Notes łączy się z bazą danych Notes i tworzy streszczenia dokumentów Notes i załączników. Streszczenie jest plikiem w formacie XML, jeden na każdy streszczony dokument, który zawiera obiekt i pełny tekst.

### Wymagania wstępne

Przed uruchomieniem programu IBM Web Crawler dla Notes wymagane jest zainstalowanie:

- Lotus Notes wersja 5.0.5 lub nowsza.
- PKZIP wersja 2.50, jeżeli mają być obsługiwane załączniki, które są samorozpakowującymi się archiwami ZIP.

### Wykonanie testowego odwiedzenia

Wybierz kolejno polecenia **Start** → **Programy** → **Wiersz polecenia**. Przejdź do katalogu, w którym zainstalowany jest program IBM Web Crawler dla Notes, a następnie przejdź do katalogu notes-run. Na przykład:

```
cd c:\<katalog_instalacyjny>\gcs\notes-run
```

Odwiedzanie Notes jest kontrolowane przez dwa pliki:

- Listę źródeł, którą należy zmodyfikować, aby określić, które bazy danych Notes mają być odwiedzone. W pliku tym podaje się nazwy serwerów Notes, adresy IP, nazwy plików .nsf itp. Na przykład lista źródeł może zawierać 34 bazy danych Notes; to, która zostanie odwiedzona, zależy od dalszej konfiguracji opisanej poniżej. Lista źródłowa może być plikiem XML albo bazą danych Notes (plikiem .nsf).
- Plik konfiguracyjny określa listę źródeł, które źródła mają być odwiedzane, jakiego rodzaju załączniki mają być przetworzone, formaty wyjściowe itp. Plik konfiguracyjny jest zawsze plikiem XML.

Aby zweryfikować, czy program IBM Web Crawler jest poprawnie zainstalowany, należy odwiedzić jego testową bazę danych. Używając edytora tekstu należy upewnić się, czy lista źródłowa testSources.xml zawiera poprawną ścieżkę i nazwę pliku do bazy danych test.nsf, która umieszczona jest w podkatalogu notes-run programu IBM Web Crawler. Najpierw należy wykonać kopię zapasową oryginalnych plików. **Zalecenie:** *Edytuj plik ostrożnie:* każdy błąd może spowodować niewłaściwe działanie programu IBM Web Crawler. Następnie należy zapisać zmiany.

Potem należy przetestować instalację, odwiedzając załączoną testową bazę danych test.nsf. Wpisz: crawlNotes crawlTestXml

Plik crawlNotes.bat uruchamia program IBM Web Crawler z plikiem crawlTestXml określonym jako plik konfiguracyjny; rozszerzenie .xml jest automatycznie dodawane do nazwy pliku konfiguracyjnego. Program IBM Web Crawler powinien zaraportować odwiedzenie i streszczenie dwóch dokumentów, każdy z jednym załącznikiem.

Po zakończeniu pracy programu IBM Web Crawler można przejrzeć streszczenia oraz pliki protokołów umieszczone w katalogach określonych w pliku konfiguracyjnym.

### Konfigurowanie własnego odwiedzenia Notes

Po pomyślnym przetestowaniu można przystąpić do rzeczywistej konfiguracji programu.

1. Utwórz listę źródłowych baz danych. Dopisz do pliku bazy danych Notes, które mają być odwiedzane.

Aby zidentyfikować bazy danych Notes, które mają być odwiedzane, zmodyfikuj plik testSources.xml albo użyj klienta Notes do utworzenia i zaktualizowania bazy danych testSources.nsf. Parametry, które mogą być ustawione w plikach źródłowych zostały objaśnione w sekcji Parametry listy źródeł.

2. Ustaw konfigurację programu odwiedzającego. Plik konfiguracyjny programu jest plikiem tekstowym w formacie XML.
  - Jeżeli źródła są umieszczone w pliku XML, zmodyfikuj plik `crawlTestXml.xml` i ustaw parametr `sourcesInXmlFile`, aby wskazywał na pliki źródłowe.
  - Jeżeli źródła umieszczone są w bazie danych Notes, zmodyfikuj plik `crawlTestNsf.xml` i ustaw parametr `sourcesInNotesDB`, aby wskazywał na bazę danych źródeł. "Plik konfiguracyjny programu IBM Web Crawler" na stronie 77 zawiera listę parametrów, które można ustawić.

Po zakończeniu konfiguracji listy źródeł, uruchom program IBM Web Crawler, wpisując: `crawlNotes plik_konfiguracyjny`.

Jeżeli lista źródeł umieszczona jest w bazie danych Notes, uruchom program IBM Web Crawler, jak na przykładzie: `crawlNotes crawlTestNsf`

Po zakończeniu pracy programu IBM Web Crawler można przejrzeć streszczenia oraz pliki protokołów, umieszczone w katalogach określonych w pliku konfiguracyjnym.

## Parametry listy źródeł

Lista źródeł zawiera opisy baz danych Notes, które mają być odwiedzone. Listy źródeł w formacie pliku `plik.xml` zawierają element `notesDataSources` z jednym lub kilkoma elementami `oneDBInfo`. Każdy element `oneDBInfo` zawiera:

**id**            Liczbowy identyfikator bazy danych. Jest on używany przez parametr `range` w pliku konfiguracyjnym.

### **serverName**

Nazwa serwera, na którym umieszczona jest baza danych. Aby określić lokalną bazę danych, wystarczy podać pusty tekst "".

### **pathAndFileName**

Pełna ścieżka i nazwa pliku bazy danych na serwerze. Ścieżkę i nazwę pliku należy zakończyć rozszerzeniem `.nsf`.

### **viewName**

Nazwa widoku Notes odwiedzanej bazy danych.

### **ipAddress**

Opcjonalne. Adres IP serwera; jeżeli jest podany, nie będzie używany serwer DNS. Opcja ta jest przydatna w sytuacji, w której serwer DNS nie jest w stanie określić adresu IP na podstawie nazwy serwera. W systemie Windows adres IP może być określony za pomocą komendy `nslookup nazwa_serwera`.

### **dateLastCrawled**

Opcjonalne. Data ostatniego odwiedzenia bazy danych. Wartość ta zostanie automatycznie zmieniona, za wyjątkiem sytuacji, w której w pliku konfiguracyjnym parametr `update Date Last Crawled` będzie ustawiony na `'no'`.

**tries**        Opcjonalne. Liczba prób odwiedzania bazy danych, jeżeli odwiedzenie nie było pomyślne (przekroczono limit czasu).

### **fieldSubstitutions**

Odwzorowania określają sposób zastępowania pól bazy danych Notes w wyjściowym pliku XML. Może ono zawierać jeden lub kilka elementów podstawienia, każdy z dwoma atrybutami:

- **Original**: nazwa pola, które zostanie zastąpione w wyjściowym dokumencie XML, jeśli istnieje.
- **Replace**: nowa nazwa pola, która zastępuje poprzednią nazwę w wyjściowym pliku XML.



Lista źródeł w bazie danych Notes może być sprawdzona i poprawiona za pomocą klienta Notes. Uruchom klienta Notes i wybierz kolejno polecenia **Plik** → **Baza danych** → **Otwórz**. Kliknij **Przeglądaj**, aby odszukać bazę danych testSources.nsf w katalogu x:\<install directory>/gcs/notes-run. Testowa baza danych, dostarczana z programem, również może być zmieniana za pomocą klienta Notes. Uruchom klienta Notes i wybierz kolejno polecenia **Plik** → **Baza danych** → **Otwórz**. Kliknij **Przeglądaj**, aby odszukać bazę danych test.nsf w katalogu x:\<install directory>/gcs/notes-run.

## Parametry pliku konfiguracyjnego

Poniżej opisano parametry, które mogą być zmieniane w pliku konfiguracyjnym. Parametry, dla których podano wartości domyślne, można pominąć.

Plik sourcesInXml lub element sourcesInNotesDB określają listę źródeł. Źródła w formacie XML lub w postaci bazy danych Notes mogą zawierać informacje identyfikujące bazę danych do odwiedzenia.

Element runInfo zawiera parametry kontrolujące jedno uruchomienie. Poniższe parametry mają zastosowanie do wszystkich odwiedzanych baz danych Notes w podanym użyciu programu odwiedzającego:

### rangeSpecify

Są to identyfikatory baz danych do odwiedzenia. Identyfikatory są liczbami określonymi w polu id listy źródeł. Parametr może być określony jako lista (oddzielona przecinkami) elementów lub zakresów elementów, np. 1-4, 15, 25-31.

### SummaryDirectory

Określa katalog główny dla tworzonych streszczeń. Streszczenia są zapisywane do podkatalogów w tym katalogu.

### MaxThreads

Określa liczbę jednoczesnych wątków odwiedzających. Każda baza danych Notes jest przeglądana przez jeden wątek. W danej chwili przeglądanych jest wiele baz danych.

### doIncrementalCrawl

Wartość domyślna: 'no'. Jeżeli jest ustawiona na 'yes', przetwarzane są tylko dokumenty utworzone/zmienione od 'summarizeThisDateAndLater'. Jeżeli parametr 'summarizeThisDateAndLater' nie jest określony, program odwiedzający użyje pola bazy danych DateLastCrawled określonego w liście źródeł bazy danych. Jeżeli doIncrementalCrawl=no, przetwarzane będą wszystkie dokumenty, niezależnie od daty.

### summarizeThisDateAndLater

Format tego pola jest następujący: MM/dd/yyyy gg:mm a sc, na przykład 01/01/2000 01:11 PM PDT. Jeżeli data i godzina nie zostaną określone, streszczane będą wszystkie dokumenty od daty ostatnich odwiedzin określonej w liście źródeł (jeżeli doIncrementalCrawl jest ustawione na 'yes') lub niezależnie od daty (jeżeli doIncrementalCrawl jest ustawione na 'no').

### detachAttachments

Wartość domyślna: 'yes'. Oznacza ona odłączanie i streszczanie załączników. Rodzaje plików załączników do przetwarzania są umieszczone w pliku konfiguracyjnym. W przeciwnym wypadku załączniki są ignorowane.

### attachmentFilenameFormat

Wartość domyślna: l (długi). Może też być: s (krótki). Długie nazwy plików kodują typ, serwer, nazwę bazy danych, Notes itp. W krótkich nazwach kodowany jest typ oraz Notes id.

**processAttachmentsAfterwards**

Wartość domyślna: 'no'. W przeciwnym wypadku pliki załączników nie są streszczane podczas streszczania bazy danych Notes. Zamiast tego zapisywany jest rekord `notesCrawl-attachments.bat` dla każdego załącznika, określający komendy służące do wykonania streszczenia pliku. Użytkownik może samodzielnie napisać i uruchomić skrypt, który później wykona streszczanie i usuwanie załączników. Taki sposób przetwarzania wymaga znacznej ilości miejsca na dysku.

**saveAttachmentFiles**

Wartość domyślna: 'no'. W przeciwnym wypadku oryginalne pliki nie są usuwane po zakończeniu przetwarzania. Opcja ta jest poprawna, jeżeli `processAttachmentsAfterwards` jest ustawione na 'no'. Jeżeli `processAttachmentsAfterwards` jest ustawione na 'yes', potrzebne będzie miejsce na dysku do przechowywania zapisanych załączników.

**MaximumNumberOfDetachingErrors**

Wartość domyślna: 10. Jest to maksymalna liczba błędów w przetwarzaniu załączników, np. brak miejsca na dysku, które program odwiedzający będzie tolerować, zanim przerwie odwiedzanie.

**saveURLsToFile**

Wartością domyślną jest 'no'. Jeśli zostanie określona, znalezione w elementach dokumentu Notes adresy URL zostaną zapisane w pliku o następującej nazwie: `nazwa_bazy_danych(bez ścieżki i rozszerzenia .nsf) + ".html"`.

**updateDateLastCrawled**

Wartość domyślna: 'yes'. Jeśli ustawiona jest na 'no', wartość `dateLastCrawled` w pliku źródeł nie jest aktualizowana.

**tempDirectory**

Wartość domyślna: `c:\temp`. Katalog ten jest używany do zapisywania plików tymczasowych.

**logSummaryDirectory**

Wartość domyślna: 'log'. Jest to katalog, w którym zapisywane są pliki protokołów.

**loggerPriority**

Wartość domyślna: 'info'. To ustawienie definiuje priorytet programu protokołującego. Możliwe ustawienia (od najbardziej istotnego): `error`, `warn` lub `info`. Na przykład, jeżeli priorytet protokołowania zostanie określony jako 'warn', będą protokołowane komunikaty z poziomem 'warn' i 'error'.

Element `attachments` zawierający elementy `include` identyfikuje rozszerzenia plików, które mają być przetwarzane.prz.

## Wykluczanie programu IBM Web Crawler z serwera

Ze względu na ochronę i wydajność administrator programu EIP może uniemożliwić odwiedzenie określonych serwerów lub stron przez instancje nawigatorów. Może zaistnieć konieczność ograniczenia aktywności nawigatorów na określonych serwerach i stronach.

Administratorzy serwerów mogą nakazać programowi IBM Web Crawler pomijanie serwerów lub stron przy użyciu pliku *strategii dostępu*. Plik ten jest tworzony w oparciu o wskazówki umieszczone w dokumencie *A Standard for Robot Exclusion (Standard wyłączenia nawigatorów)* (patrz <http://info.webcrawler.com/mak/projects/robots/norobots.html>).

- Przed rozpoczęciem nawigowania, jak również okresowo w późniejszym czasie, program IBM Web Crawler sprawdza plik reguł dostępu `http://adres_serwera/robots.txt`.
- Plik składa się z wierszy o postaci



*pole*:<opcjonalna\_spacja>*wartość*<opcjonalna\_spacja>

Jeżeli *polem* jest pole User-Agent, a *wartość* to IBM-WebCrawler lub \* , następujące po nich wiersze Disallow (aż do kolejnego wiersza User-Agent) określają częściowe adresy URL, które mają być pomijane. Dowolny adres zaczynający się od podanej wartości (pełnej lub częściowej ścieżki) nie będzie odwiedzany.

Na przykład:

Disallow: /help

powoduje pomijanie pliku /help.html i pliku /help/index.html.

Disallow: /help/

powoduje pomijanie pliku /help/index.html, ale umożliwia pobranie pliku /help.html.

Brak wartości umożliwia odwiedzanie wszystkich adresów URL.

- Wiersze można oddzielać pustymi wierszami.
- Wpisując znak #, można dodawać komentarze. Część wiersza występująca po tym znaku jest uważana za komentarz.

Oto kilka przykładów:

- Ten plik /robots.txt powoduje pomijanie danego serwera przez wszystkie roboty.

```
zabroń wszystkim
```

```
User-agent: *
```

```
Disallow: /
```

- Ten plik /robots.txt umożliwia odwiedzanie tego serwera tylko programowi IBM Web Crawler i nie wprowadza dla niego żadnych ograniczeń.

```
zezwól tylko programowi IBM
```

```
User-agent: *
```

```
Disallow: /
```

```
User-agent: IBM-WebCrawler
```

```
Disallow: # bez ograniczeń
```

- Ten plik /robots.txt określa, że wszystkie roboty mają pomijać adresy URL w podfolderach temp, development i testing folderu htmldocs, że agent użytkownika o nazwie IBM-WebCrawler ma dostęp do podfolderów development i testing (może w nich nawigować) oraz że roboty xyz i wxyz mają całkowicie pomijać serwer.

```
bardziej realistyczny przykład
```

```
User-agent: *
```

```
Disallow: /htmldocs/temp
```

```
Disallow: /htmldocs/development
```

```
Disallow: /htmldocs/testing
```

```
User-agent: xyz
```

```
User-agent: wxyz
```

```
Disallow: /
```

```
User-agent: IBM-WebCrawler
```

```
Disallow: /htmldocs/temp
```



---

## Rozdział 6. Wprowadzenie do sterowania przepływem pracy

Program obsługi przepływu pracy EIP umożliwia sterowanie przepływem i wydajnością pracy w firmie. Użytkownicy przetwarzający dane otrzymane w wyniku wykonywania wyszukiwań stowarzyszonych muszą często podejmować decyzje dotyczące działań, które mają być wykonane. Przepływ pracy umożliwia określenie z wyprzedzeniem, w jaki sposób użytkownicy mają wykonać określone zadanie.

Przepływ pracy można zautomatyzować, konfigurując profile i reguły sterujące działaniem składników przepływu pracy. Można również określać poziom ograniczeń w systemie, kontrolując dostęp i uprawnienia użytkowników za pośrednictwem zestawów uprawnień i list kontroli dostępu.

---

### Pojęcie przepływu pracy

Większość czynności biznesowych można określić jako zestaw powiązanych ze sobą procesów. Praca jest przekazywana między poszczególnymi pracownikami i działami przedsiębiorstwa. Niektóre proste procesy składają się tylko z kilku etapów, podczas gdy bardziej złożone wymagają współpracy wielu pracowników w różnych działach przedsiębiorstwa.

Przepływ pracy umożliwia przenoszenie pracy za pośrednictwem procesu i podejmowanie decyzji dotyczących prac w trakcie realizacji tego procesu. Na przykład firma Ubezpieczenia XYZ otrzymuje pocztą duże ilości formularzy dotyczących odszkodowań. W procesie sprawdzania wniosków o odszkodowania likwidatorzy szkód gromadzą dokumenty, takie jak fotografie, wyceny lub raporty rzeczoznawców. Pracownicy spędzają wiele godzin dziennie otwierając, sortując, wypełniając i sprawdzając formularze, jak również gromadząc dokumenty wymagane do ostatecznego zatwierdzenia wniosku.

Po odebraniu i sprawdzeniu tych informacji, pracownicy przekazują je kolejnym pracownikom. W procesie kompletowania pakietu dokumentów, związanych z wnioskiem o odszkodowanie, operacje na tym pakiecie mogą wykonywać pracownicy wielu działów.

---

### Jak korzystać z funkcji przepływu pracy

Podobnie jak w przypadku firmy Ubezpieczenia XYZ, większość przedsiębiorstw przetwarzających dokumenty wykonuje niektóre lub wszystkie spośród poniższych zadań:

- Archiwizowanie dokumentów w celu ich późniejszego wykorzystania.
- Gromadzenie dokumentów, formularzy i informacji z różnych źródeł, a następnie przekazywanie ich do dalszego przetworzenia.
- Sprawdzanie zgodności przychodzącej poczty z przetwarzanymi aktualnie dokumentami.

*Przepływ pracy* reprezentuje sekwencję czynności wymaganą do wykonania pracy. Określa on czynności, które mogą zostać wykonane na grupie dokumentów lub danych oraz ścieżkę, którą ta grupa dokumentów musi przejść w procesie przepływu pracy. Przepływ pracy odzwierciedla sposób wykonywania pracy i precyzyjnie definiuje jej zakres i ograniczenia. Jest to definicja sekwencji czynności i zadań oraz połączeń i związków między nimi. Przepływ pracy określa kryteria stosowane przy podejmowaniu decyzji dotyczących kolejnych etapów pracy. Informacje o tworzeniu przepływu pracy zawiera publikacja *Workstation Application Programming Guide*. Informacje o używaniu klienta z przepływem pracy zawiera podręcznik *Installing, Configuring, and Managing the eClient*.

## Synchronizowanie identyfikatorów użytkowników i grup

Niniejsza sekcja przedstawia sposób synchronizowania identyfikatorów użytkowników i grup między produktami EIP, Content Manager i MQ Series Workflow.

Zarządzając użytkownikami w systemach Content Manager lub Enterprise Information Portal można również zarządzać użytkownikami systemu MQSeries Workflow. Dlatego przy tworzeniu, modyfikowaniu lub usuwaniu identyfikatorów użytkowników i grup w systemach Content Manager lub Enterprise Information Portal, należy tę operację wykonać również dla serwera MQSeries Workflow.

Ponieważ produkty Content Manager i Enterprise Information Portal współużytkują identyfikatory użytkowników i grup, po utworzeniu użytkownika lub grupy przy użyciu klienta administracyjnego systemu z włączoną funkcją przepływu pracy, jeśli serwer MQSeries Workflow nie jest uruchomiony, zostanie wyświetlony błąd informujący, że użytkownik lub grupa nie zostali utworzeni na serwerze MQSeries Workflow. Identyfikator użytkownika nadal istnieje w systemie Content Manager lub Enterprise Information Portal, ale nie będzie go na serwerze MQSeries Workflow.

Aby zsynchronizować identyfikatory użytkowników i grup używane w systemie zarządzania danymi, należy uruchomić program synchronizacyjny przepływu pracy. Jeżeli serwer MQSeries Workflow został zainstalowany samodzielnie, sprawdź, czy jest on uruchomiony. Otwórz monitor usług i sprawdź status usługi MQSeries Workflow *X.X* - FMC (gdzie *X.X* jest wersją zainstalowanego przepływu pracy). Jeśli usługa MQSeries Workflow nie została włączona, uruchom ją lub przejdź do katalogu WFInstall i uruchom plik o nazwie **CMBWFStart.bat**. Wykonaj następujące czynności, aby uruchomić program synchronizujący:

1. Przejdź do katalogu, w którym jest zainstalowany produkt Enterprise Information Portal. Domyślnym katalogiem jest **C:\CMBROOT**.
2. Wpisz **EIPUser2WF.bat**.  
**Ważne:** Jeśli używasz zdalnego serwera przepływu pracy, musisz przed wpisaniem **EIPUser2WF.bat** uruchomić pliki skryptów serwera RMI (**cmbsvregist81.bat** lub **cmbsvregist81.sh**). Plik **EIPUser2WF.bat** używa także informacji z dwóch plików INI do odnalezienia lokalnego albo zdalnego serwera: **cmbsvcs.ini** i **cmbsvcclient.ini**. Sprawdź, czy w pliku **cmbsvcs.ini** jest wpisana wartość **LOCAL**, jeśli używasz lokalnego serwera lub **REMOTE**, jeśli używasz zdalnego serwera. Plik **cmbsvcclient.ini** zawiera informacje na temat położenia serwera.
3. Wpisz wymagane informacje dla nazwy bazy danych EIP, identyfikatora użytkownika, hasła i schematu.

Po wpisaniu wymaganych informacji program synchronizacyjny skopiuje wszystkich użytkowników systemów Content Manager lub Enterprise Information Portal do serwera MQSeries Workflow Server. Po zakończeniu jego działania nie będzie się już pojawiać błąd informujący, że dany użytkownik albo grupa nie są zdefiniowani w serwerze MQSeries Workflow.

Podczas usuwania identyfikatorów użytkowników lub grup w systemach Content Manager lub Enterprise Information Portal należy tę operację wykonać również dla serwera MQSeries Workflow. Jeśli identyfikator użytkownika albo grupy nie istnieje w serwerze MQSeries Workflow, nie można go usunąć z produktów Content Manager ani Enterprise Information Portal. Na przykład możesz utworzyć identyfikator użytkownika lub grupy w programie Content Manager lub Enterprise Information Portal przy wyłączonym serwerze przepływu pracy EIP. Następnie możesz włączyć opcję przepływu pracy przy próbie usunięcia ID użytkownika albo grupy. Ponieważ ten identyfikator użytkownika lub grupy nie występuje w przepływie pracy MQSeries, zostanie wyświetlony odpowiedni komunikat o błędzie. Aby

usunąć błąd, należy uruchomić program narzędziowy EIPUser2WF, aby zsynchronizować identyfikatory użytkowników i grup, a następnie usunąć wybrany identyfikator.

## Ponowne instalowanie serwera EIP z włączoną opcją przepływu pracy

Jeśli na serwerze EIP został włączony produkt EIP Workflow, a zamierzasz zainstalować nowy serwer EIP, musisz usunąć z serwera MQSeries Workflow wszystkie dane produktu EIP Workflow.

Przed usunięciem serwera EIP wykonaj następujące czynności, aby usunąć wszystkie dane produktu EIP z serwera MQSeries Workflow. **Ważne:** Czynności te należy wykonywać w podanej poniżej kolejności.

1. Zatrzymaj wszystkie instancje przepływu pracy, używając programu eClient lub funkcji API przepływu pracy. Zatrzymanie instancji spowoduje usunięcie wszystkich pozycji pracy.
2. Usuń listy prac i szablony przepływu pracy przy użyciu klienta administracyjnego EIP lub odpowiednich funkcji API.
3. Usuń wszystkich użytkowników i grupy z serwera EIP w chwili, gdy jest włączony przepływ pracy, aby z serwera MQSeries Workflow także zostali usunięci użytkownicy i grupy.
4. Zrekonfiguruj serwer MQSeries Workflow do współpracy z EIP:
  - a. Włóż do napędu CD-ROM płytę z produktem EIP. Przejdź w wierszu komend do katalogu WfInstall .
  - b. W wierszu komend wpisz:

```
fmcibie -iCMBWFAdmin.fdl -uadmin -ppassword -o -f
```

**Ważne:** Jeśli usuniesz bazę danych EIP przed usunięciem identyfikatorów użytkowników i grup z produktów EIP i MQSeries Workflow, a następnie spróbujesz utworzyć takie same identyfikatory użytkowników i hasła, jakie istniały w usuniętej bazie danych EIP, zostanie zgłoszony błąd informujący, że nie można dodawać użytkowników ani grup do EIP. Komunikat ten brzmi: DGL2616A: Dodanie użytkownika nie powiodło się: XXX -DGL2485A: Ten użytkownik przepływu pracy już istnieje. Wykonaj następujące czynności, aby usunąć ten problem:

1. Wyłącz usługę przepływu pracy EIP w kliencie administracyjnym tego systemu.
2. Wyloguj się z klienta administracyjnego EIP i zaloguj do niego ponownie. Jeśli zalogujesz się po wyłączeniu usługi przepływu pracy, możesz utworzyć te same identyfikatory użytkowników i grup, które istniały w systemie EIP i nadal istnieją w serwerze MQSeries Workflow.
3. Włącz usługę przepływu pracy po utworzeniu identyfikatorów użytkowników i grup, które już istnieją w serwerze MQSeries Workflow.

## Aktualizowanie identyfikatorów użytkowników i grup między MQSeries Workflow i bazą danych EIP

Dla każdego identyfikatora użytkownika lub grupy, który istnieje w serwerze MQSeries Workflow i nie istnieje w bazie danych EIP, należy wykonać podane poniżej czynności w celu zsynchronizowania identyfikatorów i grup.

1. Utwórz plik i wpisz następujący tekst:
  - CODEPAGE 1252
  - FM\_RELEASE V3R3 2
  - DELETE PERSON 'User1'
  - DELETE PERSON 'User2'

- DELETE PERSON 'User3'

gdzie User1, User2 i User3 są użytkownikami, których chcesz usunąć. Możesz podać dowolną potrzebną liczbę użytkowników.

2. Zapisz plik i uruchom w wierszu komend następującą komendę:

```
fmcibie -u admin -ppassword -i DeletePersons.fdl -f -o
```

gdzie DeletePersons.fdl jest nazwą pliku utworzonego w poprzednim kroku.

Dla każdego identyfikatora użytkownika lub grupy, który istnieje w serwerze MQSeries Workflow i nie istnieje w bazie danych EIP, należy wykonać podane poniżej czynności:

1. Utwórz plik tekstowy i wpisz następujący tekst:

- CODEPAGE 1252
- FM\_RELEASE V3R3 2
- DELETE ROLE 'Group1'
- DELETE ROLE 'Group2'
- DELETE ROLE 'Group3'

gdzie Group1, Group2 i Group3 są grupami, które chcesz usunąć. Możesz podać dowolną potrzebną liczbę grup.

2. Zapisz plik i uruchom w wierszu komend następującą komendę:

```
fmcibie -u admin -p password -i DeleteGroups.fdl -f -o
```

gdzie DeleteGroups.fdl jest nazwą pliku utworzonego w poprzednim kroku.

Podczas usuwania identyfikatora użytkownika z produktu EIP i przy włączonej obsłudze przepływu pracy może się pojawić następujący błąd:

Usunięcie użytkownika nie powiodło się [RC=12]

Jeśli usunięcie użytkownika nie powiedzie się, system wygeneruje plik temp.log, który zawiera szczegółowe informacje na temat błędu. Plik ten jest zapisywany w katalogu x:\CMBROOT . **Wskazówka:** Jeśli próbujesz usunąć identyfikator użytkownika, który służył do zalogowania się do klienta przepływu pracy MQSeries, operacja usunięcia zostanie zakończona niepowodzeniem.

---

## Planowanie przepływu pracy

Przed rozpoczęciem definiowania przepływu pracy należy przeanalizować rodzaj prac wykonywanych w organizacji, sposób i miejsce ich wykonywania oraz osoby odpowiedzialne za ich wykonanie. Ten etap planowania wykonuje administrator lub analityk biznesowy.

Co jest produktem końcowym? Produkt końcowy może być wynikiem całej pracy wykonanej w organizacji, w jej jednym dziale lub przez określonych pracowników z różnych działów. Na przykład produktem końcowym procesu przetwarzania wniosków o odszkodowania w firmie Ubezpieczenia XYZ jest list wysyłany do posiadacza polisy, powiadamiający o przyjęciu lub odrzuceniu danego roszczenia.

Należy przeanalizować informacje, które muszą zostać przetworzone w celu uzyskania produktu końcowego, określić wymagane czynności i miejsce ich realizacji oraz podjąć decyzję, w jaki sposób w procesie przepływu pracy mają być przesyłane informacje.

## Informacje do przetworzenia

Informacje, które muszą zostać przetworzone przez użytkowników w danym przedsiębiorstwie. Jakiego typu danych wejściowych wymaga produkt końcowy? Jakie dokumenty trzeba przetworzyć?

*Element pracy* nie jest bieżącym dokumentem. *Element pracy* zawiera odniesienia do dokumentu oraz inne informacje na jego temat, takie jak status, data utworzenia itd. *Pozycję zadania* może stanowić dowolny element (dokument lub obiekt) pochodzący z serwera danych. Na przykład firma Ubezpieczenia XYZ otrzymuje najpierw formularze wniosków o odszkodowanie, a dopiero później związane z nimi dokumenty, takie jak fotografie, wyceny i ekspertyzy rzeczoznawców.

## W jaki sposób informacje są przetwarzane

Kto może najlepiej wykonać dany etap procesu? Pracownik administracyjny może na przykład sprawdzić prawidłowość informacji podanych na formularzu roszczenia, a następnie zarejestrować wniosek do czasu otrzymania określonego dokumentu od posiadacza polisy. Po odebraniu dokumentu likwidator szkód może być odpowiedzialny za odnalezienie formularza wniosku, którego dotyczy ten dokument, i zatwierdzenie tego dokumentu.

*Lista prac* przypomina kolejkę prac utworzoną dla jednego lub wielu pracowników. Formularze wniosków mogą być grupowane na *liście prac* udostępnionej dla określonych likwidatorów szkód, którzy zajmują się roszczeniami. Lista prac to odfiltrowany widok elementów pracy. Pracownicy widzą tylko te pozycje na liście prac, które zostały im udostępnione.

Definicja list prac służy do filtrowania elementów pracy w celu obsługi poszczególnych fragmentów procesu przetwarzania roszczeń, takich jak gromadzenie fotografii, wycen lub raportów. Lista prac może składać się z prac pochodzących z innych przepływów pracy. Na przykład lista prac przeznaczona dla jednego likwidatora szkód może zawierać wyceny związane z jednym roszczeniem, fotografie związane z drugim roszczeniem i raporty rzeczoznawców dotyczące trzeciego roszczenia. Czynności wykonywane przez likwidatora szkód dla każdego elementu na danej liście prac mogą się różnić. Likwidator może na przykład przejrzeć wycenę i zatwierdzić pierwszy wniosek. Może się zdarzyć, że przed rozpoczęciem wykonywania czynności związanych z fotografiami zaistnieje konieczność poczekania na większą liczbę informacji dotyczących drugiego roszczenia. W wypadku trzeciego roszczenia likwidator może przesłać raporty rzeczoznawców do innego pracownika, który zajmie się realizacją kolejnych czynności związanych z obsługą tego roszczenia.

## Czynności, które należy wykonać

Należy rozważyć czynności, które trzeba wykonać na danych w przepływie pracy. Pracownik zajmujący się roszczeniami może zaakceptować formularz roszczenia lub odrzucić go jako niekompletny. *Lista czynności* określa czynności, które użytkownik może wykonać w pracy.

Na przykład w zależności od tego, czy roszczenie spełnia początkowe warunki, pracownik może wybrać opcję kontynuowania przetwarzania formularza w przepływie pracy lub opcję odrzucenia formularza.

## Przepływ informacji w procesie

Należy rozważyć, jak mają przepływać informacje i czynności. Na przykład, kiedy ma być przeglądany początkowy formularz roszczenia? Jakie dodatkowe dokumenty są potrzebne do wykonania następnego etapu w procesie? Które kryteria określają, czy roszczenie jest akceptowane czy odrzucane? Ten przepływ informacji stanowi podstawę przepływu pracy.



Proces przepływu pracy składa się ze ścieżek określających kolejne etapy przetwarzania danych. Gdzie rozpoczyna się proces? Przepływ pracy musi mieć punkt początkowy. W firmie ubezpieczeniowej dokumentem uruchamiającym przepływ pracy jest formularz roszczenia przesyłany przez posiadacza polisy.

Po otrzymaniu wszystkich dokumentów pakiet pracy może odbyć dalszą drogę aż do czynności końcowej — na przykład przyjęcia wniosku o odszkodowanie.

## Jak wszystkie elementy działają razem

Po wykonaniu analizy przetwarzanych informacji, określeniu czynności, które będą wykonywane i podjęciu decyzji o przepływie informacji użytkownik może utworzyć diagram przepływu pracy, czyli graficzną reprezentację przepływu pracy. Do utworzenia diagramu można użyć programu budującego przepływ pracy.

Diagram przepływu pracy przedstawia sposób przekazywania pracy między różnymi czynnościami w procesie oraz zadania wymagane do wykonania poszczególnych czynności. Opisuje on przepływ, główne elementy i kluczowe punkty przepływu pracy.

Każdy symbol na diagramie przepływu pracy reprezentuje punkt, w którym wykonywana jest praca. Roszczenie ubezpieczeniowe musi być zbadane. Trzeba zebrać dokumentację uzupełniającą, a następnie zatwierdzić roszczenie lub odrzucić je w zależności od określonych kryteriów. Więcej informacji na temat symboli określających procesy w programie budującym przepływ pracy można znaleźć w sekcji “Tworzenie przepływu pracy” na stronie 102.

---

## Korzystanie ze składników przepływu pracy programu Enterprise Information Portal

W tej sekcji opisane zostały komponenty przepływu pracy. Dostęp do wszystkich komponentów możliwy jest za pomocą klienta administracyjnego. **Wskazówka:** W programie EIP w wersji 8 wprowadzono kilka zmian w przepływie pracy (między innymi w pojemniku z wersji 7.1) związanych z nową architekturą produktu Content Manager w wersji 8.

## Używanie programu budującego przepływ pracy




Program budujący przepływ pracy służy do graficznego definiowania i tworzenia przepływu pracy w grupie roboczej, w dziale lub w przedsiębiorstwie. **Ograniczenie:** proces migracji EIP przenosi użytkowników z baz danych w wersji 7.1. EIP w wersji 8.2 nie zapewnia automatycznej migracji danych przepływu pracy. Należy ponownie narysować diagramy przepływu pracy z wersji 7.1 przy użyciu programu tworzącego przepływ pracy systemu EIP wersja 8.2 i na nowo wdrożyć procesy przepływu pracy.

Przed użyciem programu budującego do utworzenia modelu przepływu pracy należy zdefiniować zbiory uprawnień, listy kontroli dostępu, użytkowników, grupy użytkowników, listy czynności i listy prac. Podczas definiowania przepływu pracy w kliencie administracyjnym użytkownik może określić domyślną dla całego przepływu pracy listę czynności. Istnieje także możliwość przypisania innej listy czynności do każdego węzła przepływu pracy. Więcej informacji na temat tych zadań można znaleźć w sekcjach “Definiowanie list czynności” na stronie 102, “Definiowanie list prac” na stronie 101 i w systemie pomocy elektronicznej.

Program budujący przepływ pracy służy do budowania przepływów pracy, ale nie można go używać do uruchamiania przepływu. Przy użyciu klienta użytkownicy wyświetlają i wykonują czynności na listach elementów prac. Więcej informacji na temat tworzenia

własnych aplikacji klienckich przeznaczonych do pracy z przepływem pracy programu Enterprise Information Portal zawiera publikacja *Workstation Application Programming Guide* i elektroniczny skorowidz interfejsu API. Tabela 7 zawiera opis trzech ikon przepływu pracy, które są wspólne dla każdego przepływu pracy. Dostępne są inne ikony przepływu pracy, takie jak podprzepływ pracy, zdarzenie, kolekcja i program zewnętrzny. Pomoc elektroniczna programu EIP opisuje szczegółowo ikony paska narzędzi.

Tabela 7. Podstawowe ikony przepływu pracy

Ikona	Opis
	Węzeł początkowy uruchamia proces przepływu pracy. Diagram procesu przepływu pracy może zawierać tylko jeden węzeł początkowy.
	Węzeł końcowy stanowi zakończenie procesu przepływu pracy. Diagram każdego nowego procesu przepływu pracy zawiera węzeł końcowy. Węzeł końcowy jest tworzony automatycznie podczas tworzenia procesu. Można go przesunąć w dowolne miejsce na diagramie. Diagram procesu przepływu pracy może zawierać tylko jeden węzeł końcowy.
	Węzeł prac stanowi powiązanie między listą prac a listą czynności w określonym punkcie procesu przepływu pracy. Węzeł prac reprezentuje punkt w procesie przepływu pracy, w którym jest wykonywane zadanie. Każdy węzeł, łącznie z węzłem początkowym i końcowym, wymaga, aby istniał zbiór czynności i użytkownik, który został przypisany do wykonania tego zbioru czynności.

## Używanie usług przepływu pracy

Program Enterprise Information Portal udostępnia usługi służące do obsługiwanego przepływających informacji. Definicje przepływu pracy i list czynności utworzone przy użyciu programu budującego przepływ pracy są przechowywane w bazie danych programu Enterprise Information Portal i w bazie danych MQSeries Workflow.

Gdy administrator systemu utworzy listę prac, informacje związane z listą zadań są trwale przechowywane w administracyjnej bazie danych. Administrator systemu może za pomocą programu EIP aktualizować, usuwać i dodawać listy prac. Gdy administrator systemu sprawdza przepływ pracy, przepływ ten jest blokowany w bazie danych programu Enterprise Information Portal i oznaczany jako wyjęty przez użytkownika, co zapobiega modyfikowaniu tego przepływu przez innych użytkowników aż do chwili jego odblokowania.

## Definiowanie list prac

Listę prac można traktować jako filtr dostępnej pracy. Lista prac to odfiltrowana lista pozycji skojarzonych z określonymi użytkownikami lub grupami użytkowników. Po zalogowaniu do programu Enterprise Information Portal użytkownicy mogą wyświetlić odfiltrowane listy pozycji zadań, które zostały im przypisane. Klient administracyjny Enterprise Information Portal służy do definiowania list prac.

Definicja listy prac składa się z reguł określających sposób prezentacji, stan i ochronę jej elementów prac. Reguły dla każdej listy prac są określane podczas jej tworzenia. W celu zarządzania dostępem do list prac należy dla listy prac utworzyć listę kontroli dostępu. Szczegółowy opis sposobu definiowania list prac można znaleźć w systemie pomocy elektronicznej. Definicja listy prac zawiera następujące elementy:

### Listy kontroli dostępu

Listy kontroli dostępu zawiera jeden lub kilka identyfikatorów użytkowników lub

grup użytkowników i związane z nimi zestawy uprawnień. Zestaw uprawnień służy do definiowania praw dostępu lub praw wykonywania określonych zadań na pracy. Lista kontroli dostępu służy do ograniczania dostępu użytkownika do elementów listy prac.

#### **Filtrowanie i sortowanie list prac**

Kryteria filtrowania i sortowania listy prac wyświetlanej dla użytkownika.

#### **Maksymalna liczba elementów na liście prac**

Maksymalna liczba elementów na liście prac.

---

## **Definiowanie list czynności**

Lista czynności zawiera wszystkie czynności, które użytkownik może wykonywać podczas pracy w przepływie pracy.

Pomoc elektroniczna klienta administracyjnego dostarcza instrukcji krok po kroku wyjaśniających proces definiowania czynności i listy czynności.

---

## **Tworzenie przepływu pracy**

Po zdefiniowaniu czynności, list czynności i list prac należy użyć programu budującego przepływ pracy systemu do utworzenia modelu przepływu pracy. Pomoc elektroniczna klienta administracyjnego dostarcza instrukcji krok po kroku wyjaśniających proces definiowania czynności i listy czynności. Program budujący przepływ pracy udostępnia wizualne wskazówki dotyczące tworzenia przepływu pracy.

---

## **Włączanie programu budującego przepływ pracy**

W tym kroku opisano uruchamianie przepływu pracy w administracyjnej bazie danych.  
**Ograniczenie:** wybrana baza danych musi znajdować się na tym samym serwerze, gdzie zainstalowano produkt MQSeries i usługi MQSeries muszą być uruchomione.

Aby włączyć przepływ pracy EIP i utworzyć definicję przepływu pracy:

1. Zaloguj się do klienta administracyjnego.
2. Jeśli istnieje wiele administracyjnych baz danych, kliknij ikonę bazy, w której chcesz włączyć przepływ pracy.
3. Kliknij Narzędzia->Usługi. Kliknij opcję Włącz przepływ pracy.
4. Wyloguj się z klienta i zaloguj ponownie. Jeśli istnieje wiele baz danych, wybierz ikonę bazy, w której przepływ pracy EIP został włączony. Zostanie wyświetlona ikona z folderem Przepływy pracy.
5. W lewym panelu głównego okna administracyjnego programu Enterprise Information Portal kliknij dwukrotnie folder **Przepływy pracy**.
6. Kliknij prawym przyciskiem myszy ikonę Definicje przepływu pracy, a następnie wybierz opcję **Nowa** w celu utworzenia definicji przepływu pracy.

**Wymaganie:** Przed zdefiniowaniem przepływu pracy należy utworzyć co najmniej jedną listę kontroli dostępu, jedną czynność i jedną listę czynności.

---

## **Uruchamianie serwera MQSeries Workflow**

Uruchom serwer MQSeries Workflow, wprowadzając w wierszu komend komendę: cmbwfsstart. Zostaną otwarte dwa okna serwera MQSeries Workflow. Okna te należy pozostawić otwarte, aby nie przerwać pracy serwera.

Jeśli usługa przepływu pracy będzie instalowana po zainstalowaniu programu Enterprise Information Portal, należy skonfigurować system Enterprise Information Portal do korzystania z tej funkcji. Konfigurację należy zmienić także, jeśli usługa przepływu pracy ma być zainstalowana na innej stacji roboczej niż klient administracyjny.

1. W oknie Administracja kliknij podzbiór **Narzędzia**.
2. Z menu wybierz opcję **Usługi**.
3. Zaznacz pole wyboru **Przepływ pracy**.
4. Po zakończeniu konfiguracji wyloguj się z klienta administracyjnego Enterprise Information Portal i zaloguj się ponownie, aby zainicjować usługę przepływu pracy. Po zalogowaniu się do klienta administracyjnego Enterprise Information Portal w lewym oknie pojawi się ikona **Definicje przepływu pracy**.

**Wskazówka:** Ikona **Definicje przepływu pracy** nie jest wyświetlana w oknie administratora, jeśli nie ma on uprawnień do administrowania przepływem pracy. Informacje na temat ograniczania dostępu użytkowników do usługi przepływu pracy można znaleźć w podręcznikach dotyczących administrowania systemami serwerów danych. Więcej informacji na temat nadawania administratorom uprawnień do zarządzania przepływem pracy można znaleźć w systemie pomocy elektronicznej.

Klient może być utworzony na podstawie przykładów i przy użyciu pakietu narzędzi dla programistów łączników.



---

## Rozdział 7. Przykładowe pliki programu IBM Web Crawler

W tym załączniku umieszczono dwa przykładowe pliki. Pierwszy plik, config-sample2.xml, zawiera przykłady parametrów konfiguracyjnych <gcs-config>. Drugi jest przykładem protokołu po zakończeniu pobierania.

---

### Przykładowy plik config-sample2.xml

W tej sekcji umieszczono przykład pliku gcs-config.

```
<!DOCTYPE gcs-config SYSTEM "config.dtd">
<gcs-config>
 <!-- Ustawienia globalne: -->
 <globals max-urls="1000000"
 num-crawlers="30"
 num-summarizers="8"
 summaries-dir="summaries"
 log-file="log/LOG.txt"
 temp-dir="temp"
 log-priority="warn"
 text-monitor="60"
 graph-monitor="2"
 connect-timeout="120"
 read-timeout="100">

 <!-- przykładowe ustawienia protokolanta -->
 <logger-config category="gcs.summaries.list.resource"
priority="info" log-file="log/resources.txt"/>
 <logger-config category="gcs.summaries.list.host" priority="info"
log-file="log/hosts.txt"/>
 <logger-config category="gcs.url.skipped" priority="info"
log-file="log/skipped_urls.txt"/>
 <logger-config category="gcs.url.fromto" priority="info"
log-layout="%m\n" log-file="log/fromto.txt"/>
 <logger-config category="gcs.http" priority="info"
log-file="log/http.txt"/>
 <logger-config category="gcs.http.connect" priority="info"
log-file="log/connecterrs.txt"/>

 <!-- określanie bazy danych
 <urlpool-config urlcontainer-class="DB2URLContainer"
urlcollection- class="DB2URLCollection">
 <urlpool-param name="dbname" value="gcs"/>
 <urlpool-param name="user" value="xxxxxx"/>
 <urlpool-param name="password" value="xxxxxx"/>
 <urlpool-param name="cachesize" value="1000"/>
 <urlpool-param name="driver"
value="COM.ibm.db2.jdbc.app.DB2Driver"/>
 </urlpool-config -->

 <!-- określanie proxy typu SOCKS
 <system-properties>
 <property name="socksProxySet" value="true"/>
 <property name="socksProxyHost" value="socks2.server.ibm.com"/>
 <property name="socksProxyPort" value="1080"/>
 </system-properties -->

 </globals>

 <group-list>
 <group name="ibm">
```

```

<crawler-config recursion-depth="-1">
 <seed-list>
 <!-- Adresy URL do rozpoczęcia pobierania: -->
 <seed url="http://gcs.stl.ibm.com/gcs/testurl.html"/>
 <seed url="http://gcs.stl.ibm.com/gcs/stl.html"/>
 <seed url="http://gcs.stl.ibm.com/gcs/ibm.html"/>
 </seed-list>

 <content-type-pattern-list>
 <!-- rozszerzenia plików w adresach URL, które nie pasują do tych wzorców,
 nie będą pobierane: -->
 <url-name-pattern name="htm*" />
 <url-name-pattern name="pdf" />
 <url-name-pattern name="gif" />
 <url-name-pattern name="zip" />
 <url-name-pattern name="txt" />
 </content-type-pattern-list>

 <include-pattern-list>
 <!-- Adresy URL, które nie pasują do tych wzorców, nie będą pobierane: -->
 <url-obj-pattern host="*.ibm.com" />
 <!-- sbo - url-obj-pattern query="*OpenDocument*" / -->
 <!-- sbo - url-obj-pattern query="*OpenView*" / -->
 <!-- url-obj-pattern query="*OpenDocument =>
OpenDocument&ExpandAll*" / -->
 <!-- url-obj-pattern query="*OpenView =>
OpenView&ExpandAll&Count=999999*" / -->
 </include-pattern-list>

 <exclude-pattern-list>
 <!-- Adresy URL, które pasują do tych wzorców, nie będą pobierane: -->
 <!-- pomijanie często używanych wzorców w naszym intranecie -->
 <url-obj-pattern file="*news*" />
 <url-obj-pattern file="*search*" />
 <url-obj-pattern file="*/afs/*" />
 <url-obj-pattern file="*/.../*" />
 <url-obj-pattern file="*bluepages*" />
 <!-- pomijanie osobistych stron WWW -->
 <url-obj-pattern file="*/~*" />
 <!-- pomijanie SOCKS: żaden adres URL nie powinien określać
tego bezpośrednio -->
 <url-regex-pattern match="*:1080/*" />
 <!-- pomijanie bram, zalecane dla zwykłych śmiertelników -->
 <url-regex-pattern match=".*[?\\=\\+\\;\\%"&].*" />
 <!-- za wyjątkiem bram zdefiniowanych poniżej... -->
 <!-- pomijam Domino -->
 <url-obj-pattern file="*.nsf*" />
 <!-- pozwalam jednak na OpenDocument -->
 <url-obj-pattern query="*OpenServer*" />
 <url-obj-pattern query="*OpenDatabase*" />
 <url-obj-pattern query="*OpenElement*" />
 <url-obj-pattern query="*OpenView*" />
 <url-obj-pattern query="*OpenAbout*" />
 <url-obj-pattern query="*OpenHelp*" />
 <url-obj-pattern query="*OpenIcon*" />
 <url-obj-pattern query="*OpenForm*" />
 <url-obj-pattern query="*OpenNavigator*" />
 <url-obj-pattern query="*OpenAgent*" />
 <url-obj-pattern query="*CreateDocument*" />
 <url-obj-pattern query="*DeleteDocument*" />
 <url-obj-pattern query="*EditDocument*" />
 <url-obj-pattern query="*SaveDocument*" />
 <url-obj-pattern query="*SearchSite*" />
 <url-obj-pattern query="*SearchView*" />
 <url-obj-pattern query="*&login*" />
 <url-obj-pattern query="*Command*" />
 <!-- Domino: unikanie permutacji OpenDocument -->

```



```

 <url-obj-pattern query="*ExpandSection*" />
 <url-obj-pattern query="*Navigate*" />
 <url-obj-pattern query="*Start*" />
 <!-- -->

 </exclude-pattern-list>
</crawler-config>

<summarizer-config>
<!-- Copy2Rdf jest domyślnym programem streszczającym. Dla tych typów użyj: -->

 <resource-handler content-type="*htm*"
 summarizable="EipHtmlSummarizable"
 summary-maker="EipHtmlRawSummaryMaker" />
 <resource-handler content-type="*pdf"
 summarizable="InsoSummarizable"
 summary-maker="InsoSummaryMaker" />

</summarizer-config>
v </group>
 </group-list>
</gcs-config>

```

---

## Analiza przykładowego pliku protokołu programu IBM Web Crawler

D:\gcs\run\log>perl loganalysis.pl log.txt

Elapsed time in Log.txt for 7710 lines was 1.84 minutes.

```

GCS was configured for 20 crawlers
 999 total crawls attempted
 137 - total crawl failures:
 21 GCSHttpConnection.ABANDONING
 12 GCSHttpConnection.CONNECT_ERROR
 16 GCSHttpConnection.UNKNOWN_HOST
 4 HTTP 403
 29 HTTP 404
 2 HTTP 500
 8 HTTP 599
 1 Read timed out
 39 Robots not allowed
 4 over max redirects
 1 unknown protocol

 862 = successfully crawled
 0 - unchanged since earlier crawl

 862 = new or changed
 468 crawled per minute

```

```

GCS was configured for 5 summarizers
 855 total summaries attempted
 0 - total summary failures:

 855 = successfully summarized
 144 gcs.summaries.list.host
 855 gcs.summaries.list.resource
 465 summarized per minute

```

```

GCS successfully crawled 134 servers to obtain 862 URL:
afqa0854.mop.ibm.com: 15
alsif1.yamato.ibm.com: 1
apache.btv.ibm.com: 1
apc.endicott.ibm.com: 2
as400service.ibm.com: 1

```

atlas.bocaraton.ibm.com: 1  
 autoproxy.ibm.com: 1  
 cer.si.ibm.com: 1  
 commerce.www.ibm.com: 1  
 crmweb.boulder.ibm.com: 3  
 d02ntcl01.ibm.com: 1  
 dacs.endicott.ibm.com: 1  
 duke.toraix.can.ibm.com: 1  
 ebcweb.austin.ibm.com: 1  
 ecspubs.ibm.com: 5  
 edaw3.fishkill.ibm.com: 1  
 endwww.endicott.ibm.com: 1  
 gcs.stl.ibm.com: 1  
 gustwick.austin.ibm.com: 1  
 ibmfnsys.somers.hqregion.ibm.com: 1  
 ibmpny1.somers.hqregion.ibm.com: 2  
 ifw-www.mul.ie.ibm.com: 1  
 iplswww.nas.ibm.com: 2  
 itirc.ibm.com: 1  
 logosite.services.ibm.com: 1  
 lt.lahulpe.ibm.com: 17  
 messaging.ibm.com: 1  
 mrsmrn04.leeds.uk.ibm.com: 1  
 online.lahulpe.ibm.com: 1  
 page.sg.ibm.com: 1  
 procure.sby1.ibm.com: 1  
 reso.somers.hqregion.ibm.com: 1  
 risc1.leipzig.de.ibm.com: 1  
 rrhhar.argentina.ibm.com: 1  
 seashore.stl.ibm.com: 1  
 secureway.raleigh.ibm.com: 15  
 service.software.ibm.com: 1  
 software.ibm.com: 1  
 techcenter.austin.ibm.com: 1  
 tr2.fishkill.ibm.com: 8080: 1  
 ucd.torolab.ibm.com: 1  
 usmweb.boulder.ibm.com: 1  
 w3-1.ibm.com: 32  
 w3-2.ibm.com: 3  
 w3-3.ibm.com: 108  
 w3-5.ibm.com: 4  
 w3.a-nz.au.ibm.com: 1  
 w3.academy.ibm.com: 1  
 w3.almaden.ibm.com: 2  
 w3.alphaworks.ibm.com: 1  
 w3.ap.ibm.com: 1  
 w3.asca.ibm.com: 7  
 w3.austin.ibm.com: 3  
 w3.boulder.ibm.com: 1  
 w3.br.ibm.com: 1  
 w3.btv.ibm.com: 1  
 w3.can.ibm.com: 40  
 w3.chq.ibm.com: 4  
 w3.coc.ibm.com: 1  
 w3.corporatetechnology.ibm.com: 1  
 w3.cupertino.ibm.com: 1  
 w3.dds.dfw.ibm.com: 17  
 w3.demopkg.ibm.com: 4  
 w3.design.ibm.com: 1  
 w3.developer.ibm.com: 3  
 w3.education.ibm.com: 1  
 w3.emea.ibm.com: 14  
 w3.enterlib.ibm.com: 7  
 w3.finsys.ibm.com: 1  
 w3.gcg.ibm.com: 1  
 w3.globalfinancing.de.ibm.com: 1  
 w3.hakozaki.ibm.com: 1

w3.houston.ibm.com: 1  
w3.hursley.ibm.com: 5  
w3.iabc.ibm.com: 1  
w3.ibm.com: 180  
w3.ibmfax.ibm.com: 1  
w3.ibmlla.ibm.com: 14  
w3.isicc.de.ibm.com: 1  
w3.itso.ibm.com: 1  
w3.japan.ibm.com: 1  
w3.knowledge.raleigh.ibm.com: 1  
w3.linux.ibm.com: 1  
w3.marketiq.ibm.com: 1  
w3.micro.ibm.com: 2  
w3.mtlisc.can.ibm.com: 1  
w3.munich.ibm.com: 1  
w3.ode.raleigh.ibm.com: 1  
w3.paylink.au.ibm.com: 1  
w3.pisc.uk.ibm.com: 1  
w3.pl.ibm.com: 1  
w3.printers.ibm.com: 1  
w3.pssc.mop.ibm.com: 1  
w3.pssed.au.ibm.com: 1  
w3.raleigh.ibm.com: 3  
w3.rchland.ibm.com: 1  
w3.research.ibm.com: 3  
w3.reserve.ibm.com: 1  
w3.rs6000.ibm.com: 1  
w3.security.ibm.com: 1  
w3.software.ibm.com: 6  
w3.ssd.ibm.com: 1  
w3.stl.ibm.com: 1  
w3.techline.ibm.com: 1  
w3.techsupp.yamato.ibm.com: 1  
w3.torolab.ibm.com: 2  
w3.usergroup.ibm.com: 1  
w3.vendor.pok.ibm.com: 1  
w3.viewblue.ibm.com: 1  
w3.watson.ibm.com: 2  
w3.wdg.uk.ibm.com: 1  
w3.ytal.yasu.ibm.com: 1  
w3.zurich.ibm.com: 1  
w3chq.disbursements.ibm.com: 1  
w3is.lagaude.ibm.com: 1  
w3md.btv.ibm.com: 1  
w3ssd.mainz.de.ibm.com: 1  
w3vm.demopkg.ibm.com: 1  
widweb.raleigh.ibm.com: 1  
wtscpok.itso.ibm.com: 1  
wwas.raleigh.ibm.com: 1  
www-1.ibm.com: 63  
www-3.ibm.com: 4  
www-4.ibm.com: 86  
www.almaden.ibm.com: 1  
www.as400.ibm.com: 1  
www.chips.ibm.com: 1  
www.ibm.com: 52  
www.ieg.ibm.com: 1  
www.patents.ibm.com: 1  
www.pc.ibm.com: 2  
www.rs6000.ibm.com: 23  
www.software.ibm.com: 9  
www.storage.ibm.com: 1  
www.watson.ibm.com: 1

GCS timed out 1 times:  
w3-3.ibm.com: 1

GCS ignored 42 URL prohibited by robots.txt:  
reso.somers.hqregion.ibm.com: 1  
w3.education.ibm.com: 1  
w3.rchland.ibm.com: 34  
w3.zurich.ibm.com: 1  
www.ibm.com: 5

GCS skipped 3846 URL (requires gcs.url logging)  
59 specified an unsupported protocol:  
protocol not supported gopher: 6  
protocol not supported mailto: 53  
1206 had content-types (lower or UPPER case, > 10) that were not included  
.2: 12  
.faq: 13  
.1: 14  
.asp: 16  
.cgi: 21  
.shtml: 90  
.pl: 92  
.gif: 157  
.nsf: 160  
.jpg: 214  
.css: 240  
516 URL were on servers and/or paths that were not included  
2065 were excluded for these reasons:  
URL longer than 254: 1  
excluded by rule 1: 1210  
excluded by rule 2: 854

---

## Rozdział 8. Używane wyszukiwania tekstowego i QBIC

Pierwsza sekcja niniejszego dodatku wyjaśnia sposób korzystania z wyszukiwania tekstowego i funkcji QBIC (Query by Image Content). Są one dostępne tylko wtedy, jeżeli przy instalacji programu EIP została wybrana wersja 7.1 programu Content Manager. W drugiej sekcji zawarte są informacje na temat ładowania prostego tekstu i danych graficznych za pomocą przykładowych aplikacji.

---

### Wyszukiwanie dokumentów za pomocą mechanizmu wyszukiwania tekstowego

Na serwerach Content Manager wersja 7.1 można zainstalować funkcję wyszukiwania tekstowego, umożliwiającą automatyczne indeksowanie, wyszukiwanie i pobieranie dokumentów przechowywanych przez program Content Manager. Dokumenty mogą być przeszukiwane przez podanie szukanych wyrazów lub całych fraz. Serwer wyszukiwania tekstowego obsługuje zarówno jednobajtowe, jak i dwubajtowe zestawy znaków oraz działa w systemie AIX i w systemie Windows.

Wyszukiwanie tekstów zawiera obsługę dokumentów strukturalnych XML, HTML i znacznikowego formatu ASCII, co umożliwia wyszukiwanie terminów w określonych sekcjach dokumentu. Można szukać danych w sekcjach zagnieżdżonych. Wyszukiwania można przeprowadzać według pełnego kontekstu XML. Można na przykład szukać słowa IBM w tytułach lub w tytule należącym do określonej sekcji. Określenie ścieżki DTD umożliwia wyszukiwanie tekstowe z dynamicznym wykorzystywaniem odpowiedniego opisu DTD każdego dokumentu, o ile tylko odniesienie do opisu DTD jest przechowywane jako metadane dokumentu.

Więcej informacji na temat planowania i instalowania systemów używających wyszukiwania tekstowego znajduje się w publikacji *Planowanie i instalowanie programu Enterprise Information Portal*.

### Włączanie serwera wyszukiwania tekstowego

Aby można było wykorzystywać serwer wyszukiwania tekstowego, trzeba włączyć jego funkcje administracyjne przed uruchomieniem klienta administracyjnego programu IBM Content Manager for Multiplatforms. Aby włączyć funkcje administracyjne:

1. Uruchom IBM Content Manager for Multiplatforms systemu serwer biblioteczny.  
Pozwól, aby serwer biblioteczny zakończył tworzenie klas indeksowych.
2. Uruchom serwer wyszukiwania na stacji roboczej, na której jest on zainstalowany, przy użyciu następującej komendy:  

```
imlss -start dlint
```

gdzie *dlint* to nazwa instancji serwera wyszukiwania tekstowego wybrana podczas jego instalowania lub przy użyciu programu narzędziowego *imlcfgsv*.

---

### Wyszukiwanie obrazów za pomocą języka QBIC

W tej sekcji umieszczono wprowadzenie do języka Query by Image Content (QBIC) i objaśniono konfigurowanie i używanie QBIC. Funkcja QBIC jest dostępna tylko wtedy, gdy zainstalowany został łącznik Content Manager wersja 7.1. QBIC jest kompatybilny z systemami operacyjnymi Windows i AIX.

## Wprowadzenie do wyszukiwania obrazów

Na serwerze wyszukiwania obrazów używana jest technika QBIC (zapytanie według zawartości obrazu) firmy IBM, która umożliwia wyszukiwanie obiektów według pewnych cech wizualnych, takich jak kolor lub tekstura. Serwer wyszukiwania obrazów analizuje je i przechowuje informacje o nich w bazie danych. Użytkownicy mogą wystosowywać zapytania o obrazy z podaniem ich wizualnych właściwości, tak ich jak kolory, tekstury i ich pozycje, nie opisując ich słowami. Zapytania dotyczące zawartości można łączyć z wyszukiwaniem tekstu i słów kluczowych, co zwiększa skuteczność wyszukiwania obrazów i danych multimedialnych.

Na każdym serwerze wyszukiwania obrazów znajduje się katalog zawierający jedną lub wiele baz danych wyszukiwania obrazów, w których są przechowywane katalogi wyszukiwania obrazów. Katalog wyszukiwania obrazów zawiera dane o wizualnych cechach zbioru obrazów. Rzeczywiste obiekty obrazów są przechowywane na serwerach obiektów w systemie IBM Content Manager for Multiplatforms. Serwer wyszukiwania obrazów działa w systemach AIX i Windows.

Więcej informacji na temat instalowania wyszukiwania obrazów znajduje się w publikacji *Planowanie i instalowanie programu Enterprise Information Portal*.

## Konfigurowanie funkcji wyszukiwania obrazów

Te instrukcje stosuje się po zainstalowaniu funkcji wyszukiwania obrazów, która jest automatycznie instalowana wraz z łącznikiem Content Manager wersja 7.1. Instalowanie wyszukiwania obrazów składa się z następujących etapów:

1. Konfigurowanie środowiska
2. Konfigurowanie serwera wyszukiwania obrazów
3. Konfigurowanie klienta wyszukiwania obrazów
4. Załadowanie przykładowych danych o obrazach

**Podczas korzystania z kreatora instalacji w systemie AIX** nie trzeba uruchamiać skryptów konfiguracyjnych bądź instalacyjnych, ani też wywoływać komend konfiguracyjnych serwera. Kreator wykonuje te zadania automatycznie.

**Podczas instalowania w systemie Windows** trzeba wykonać te czynności.

### Konfigurowanie środowiska

Zadania konfigurowania środowiska opisane w tej sekcji należy wykonać na komputerach serwera i klienta. Serwer wyszukiwania obrazów wymaga następujących zmiennych środowiska:

#### **QBICTOP**

Służy do ustalania nazw plików podczas konfigurowania wyszukiwania obrazów

#### **QbicImagePath**

Służy do ustalania nazw plików w pliku obrazu serwera

#### **QbicMaskPath**

Służy do ustalania nazw plików w pliku maski serwera

#### **QbicSketchPath**

Służy do ustalania nazw plików w pliku szkicu serwera

#### **QbicTextPath**

Służy do ustalania nazw plików w pliku tekstowym serwera

Klient wyszukiwania obrazów wymaga określenia tylko zmiennej środowiska QBICTOP.

**Przykład konfigurowania w systemie AIX:** Uruchom w systemie AIX skrypt konfiguracyjny, który spowoduje wygenerowanie skryptu instalacyjnego, a następnie skrypt instalacyjny, aby skonfigurować środowisko.

1. Uruchom następujący skrypt konfiguracyjny:

```
/usr/lpp/cmb/bin/frnconfg.iss QBICTOP
```

Parametr QBICTOP określa ścieżkę dostępu do katalogu, w którym są przechowywane pliki kontrolne (\*.ini). Zmienną QBICTOP ustaw na wartość /użytk1/cmb/qbic, gdzie parametr /użytk1 określa katalog osobisty powiązany z identyfikatorem administratora wyszukiwania obrazów. Użytkownik funkcji wyszukiwania obrazów musi mieć prawo do odczytu i zapisu w tym katalogu.

Ten skrypt generuje skrypt instalacyjny: frnsetup.iss.

2. Z katalogu osobistego użytkownika funkcji wyszukiwania obrazów uruchom skrypt:

```
./frnsetup.iss
```

Ten skrypt spowoduje rozpowszechnienie zmiennych środowiska na serwery i klienckie stacje robocze wyszukiwania obrazów.

**Przykład konfigurowania w systemie Windows:** Aby ustawić zmienne środowiska:

1. Wybierz kolejno polecenia **Start** → **Ustawienia** → **Panel sterowania**.
2. Kliknij dwukrotnie ikonę **System**.
3. Kliknij zakładkę **Środowisko**.
4. Ustaw zmienne i wartości wyświetlane w tabeli Tabela 8, wpisując je w odpowiednie pola i klikając przycisk **Ustaw**.

**Wymaganie:** Klient wyszukiwania obrazów wymaga określenia tylko zmiennej QBICTO. W środowiskach klienckich należy określić tylko zmienną QBICTOP.

*Tabela 8. Zmienne środowiska dla funkcji wyszukiwania obrazów*

Zmienna	Wartość
QBICTOP	d:\cmbroot\iss
QbicImagePath	d:\cmbroot\iss
QbicMaskPath	d:\cmbroot\iss
QbicSketchPath	d:\cmbroot\iss
QbicTextPath	d:\cmbroot\iss

Parametr d: określa dysk, na którym zainstalowano funkcję wyszukiwania obrazów.

## Konfigurowanie serwera wyszukiwania obrazów

Przed uruchomieniem serwera wyszukiwania obrazów trzeba go skonfigurować. Czynności konfiguracyjne serwera obejmują kończenie konfiguracji początkowej i weryfikowanie połączenia.

Aby skonfigurować serwer:

1. Uruchom interpreter komend, wprowadzając komendę: qbicadm
2. Wprowadź komendę **config server**. Na przykład:

```
config server LIBSRVRN FRNADMIN PASSWORD 9999
```

Parametr LIBSRVRN określa nazwę serwera bibliotecznego, FRNADMIN to identyfikator użytkownika programu Content Manager, PASSWORD to hasło programu Content Manager, a 9999 to numer portu serwera wyszukiwania obrazów.



Więcej informacji na ten temat można znaleźć w sekcji “Weryfikacja połączenia”.

## Konfigurowanie klienta wyszukiwania obrazów

Przed uruchomieniem klienta wyszukiwania obrazów lub programu administracyjnego systemu wyszukiwania należy go skonfigurować. Program administracyjny systemu Content Manager wymaga przypisania jednego aliasu. Konfigurację należy przetestować poprzez zweryfikowanie połączenia.

**Przypisywanie aliasu:** Aby można było wykorzystywać program administracyjny systemu wyszukiwania obrazów Content Manager, pracujący jako klient wyszukiwania obrazów, trzeba przypisać przynajmniej jeden alias serwera.

Aby przypisać alias:

1. Uruchom interpreter komend, wprowadzając komendę: `qbicadm`
2. Użyj komendy **add alias**. Na przykład:  
`add alias QBICSRV HOSTNAME 9999`

Parametr QBICSRV określa nazwę aliasu, HOSTNAME to nazwa hosta serwera wyszukiwania obrazów, a 9999 to numer portu serwera wyszukiwania obrazów.

### Weryfikacja połączenia:

#### Ważne:

1. Aby można było połączyć się z serwerem wyszukiwania obrazów, serwer biblioteczny musi być uruchomiony.
2. Program administracyjny systemu wyszukiwania obrazów wymaga istniejącego identyfikatora użytkownika systemu Content Manager. Aby można było pomyślnie połączyć się z serwerem bibliotecznym, identyfikatory użytkownika systemu wyszukiwania obrazów i serwera bibliotecznego muszą być takie same. Domyślną wartością dla tego identyfikatora użytkownika jest `frnadmin`. Jeżeli wartość ta zostanie zmieniona, upewnij się, że identyfikatory są zgodne.

Aby zweryfikować połączenie:

1. Po skonfigurowaniu serwera wyszukiwania obrazów i dodaniu aliasu uruchom go, wpisując w wierszu komend serwera `commsrv`.
2. Aby uruchomić interpreter komend, wpisz komendę `qbicadm`.
3. W interpreterze komend wprowadź komendę `connect`.

```
connect QBICSRV FRNADMIN
PASSWORD
```

gdzie QBICSRV to alias, FRNADMIN to identyfikator użytkownika programu Content Manager, a PASSWORD to hasło programu Content Manager.

Po pomyślnym nawiązaniu połączenia zostanie wyświetlony komunikat Library Server is LIBSRVRN.

4. Aby odłączyć serwer, wpisz komendę `disconnect`.
5. Aby zamknąć interpreter komend, wpisz komendę `quit`.

---

## Ładowanie i indeksowanie przykładowych danych

Sekcja ta zawiera objaśnienia dotyczące ładowania i indeksowania przykładowych danych tekstowych i graficznych, które są wykorzystywane w aplikacjach przykładowych. Ta sekcja ma zastosowanie tylko wtedy, gdy zainstalowany jest łącznik Content Manager wersja 7.1 wraz z opcją wyszukiwania tekstowego.

Na dysku CD programu Enterprise Information Portal znajduje się kilka przykładowych programów ładujących. W tej sekcji opisano, w jaki sposób można załadować dane graficzne i tekstowe za pomocą przykładowego programu ładującego LoadSampleTSQBICDL. Aby upewnić się, że obie funkcje programu działają poprawnie, można oddzielnie załadować dane graficzne i dane tekstowe.

### Przed załadowaniem przykładowych danych

Przed uruchomieniem programu ładującego:

1. Zaloguj się do klienta administracyjnego EIP. Kliknij kolejno opcje **Start** → **Programs** → **Enterprise Information for Multiplatforms 8.2** → **Administracja**.
2. Wybierz bazę danych i zaloguj się, używając poprawnego identyfikatora użytkownika i hasła. Jeżeli wybrana zostanie domyślna baza danych, icmnlsdb, wpisz identyfikator użytkownika **icmadmin** i hasło **password**. Jeżeli wybrana zostanie inna baza danych, wpisz odpowiednie wartości.
3. Utwórz konfigurację serwera bibliotecznego przy użyciu programu administrowania systemem Content Manager. Informacje pomocne przy realizacji tego zadania można znaleźć w systemie pomocy elektronicznej programu administrowania systemem.
4. W grupie **Dostęp** zmień właściwości dotyczące konfiguracji serwera bibliotecznego, wykonując następujące czynności:
  - a. Kliknij prawym przyciskiem nową konfigurację i wybierz opcję **Właściwości**, aby otworzyć notatnik Właściwości.
  - b. Kliknij zakładkę Dostęp.
  - c. Kliknij przełącznik **Nieograniczona liczba sesji z dowolnej stacji roboczej**.

### Tworzenie indeksu wyszukiwania tekstowego

Przed załadowaniem danych należy utworzyć pusty indeks wyszukiwania tekstowego, który może zostać użyty do indeksowania przykładowych tekstów. **Wskazówka:** indeks wyszukiwania tekstowego można utworzyć tylko na serwerze Content Manager w wersji 6.1 lub 7.1.

Aby utworzyć indeks wyszukiwania tekstowego:

1. Za pomocą następującej komendy uruchom serwer wyszukiwania na stacji roboczej, na której jest on zainstalowany:

```
imlss -start dlinst
```

gdzie dlinst to nazwa instancji serwera wyszukiwania tekstowego wybrana podczas jego instalowania lub przy użyciu programu narzędziowego imlcfgsv.
2. Uruchom program administrowania systemem Content Manager i zaloguj się do niego.
3. Z listy znajdującej się w lewym górnym okienku wybierz pozycję **Wyszukiwanie tekstowe**.
4. W lewym okienku kliknij dwukrotnie pozycję **Serwery wyszukiwania**.
5. Kliknij dwukrotnie ikonę **TM**. TM jest aliasem serwera wyszukiwania tekstowego.

6. W lewym okienku kliknij dwukrotnie folder **Indeksy**. Jeśli zostanie wyświetlony komunikat RC\_EMPTY\_LIST, należy z menu wybrać kolejno polecenia **Zaznaczone** → **Nowe**, aby utworzyć indeks.
7. W oknie Nowy indeks zdefiniuj indeks. Aby uzyskać szczegółowy opis każdego pola, kliknij przycisk **Pomoc**.

Na przykład:

**W systemie Windows:**

**Nazwa** TMINDEX

**Typ** Precyzyjny

**Pliki indeksów**

x:\cmbroot\ts\index\tminindex; gdzie x oznacza dysk instalacyjny; jeśli podana ścieżka nie istnieje, jest ona tworzona.

**Pliki robocze indeksów**

x:\cmbroot\ts\work\tminindex; gdzie x oznacza dysk instalacyjny; jeśli podana ścieżka nie istnieje, jest ona tworzona.

**Pozycja informacyjna**

Nazwa serwera bibliotecznego programu Content Manager.

Nie należy zmieniać domyślnych nazw bibliotek DLL klienta i serwera.

**W systemie AIX:**

**Nazwa** TMINDEX

**Typ** Precyzyjny

**Pliki indeksów**

/home/cltadmin/tsindex/index/tminindex; jeśli podana ścieżka nie istnieje, jest ona tworzona.

**Pliki robocze indeksów**

/home/cltadmin/tsindex/work/tminindex; jeśli podana ścieżka nie istnieje, jest ona tworzona. Należy sprawdzić, czy użytkownik ma prawo do zapisu w tym katalogu.

**Pozycja informacyjna**

Nazwa serwera bibliotecznego programu Content Manager.

8. Kliknij przycisk **OK**.
9. Kliknij dwukrotnie ikonę **TMINDEX**, aby otworzyć notatnik administracyjny TMINDEX.

## Tworzenie bazy danych, katalogu i funkcji wyszukiwania obrazów

Po utworzeniu indeksu wyszukiwania tekstowego dla przykładowych danych wyszukiwania tekstowego trzeba utworzyć bazę danych wyszukiwania obrazów i katalog dla przykładowych danych graficznych.

Aby utworzyć bazę danych, katalog i funkcje wyszukiwania obrazów:

1. Uruchom serwer wyszukiwania obrazów na stacji roboczej, na której jest on zainstalowany, wprowadzając następującą komendę:  
commsrv
2. Uruchom program administrowania systemem Content Manager i zaloguj się do niego.

3. Z listy znajdującej się w lewym górnym okienku wybierz pozycję **Wyszukiwanie obrazów**.
4. W lewym okienku kliknij pozycję **Serwery wyszukiwania obrazów**.
5. Kliknij pozycję **QBICSRV**.  
Parametr QBICSRV to nazwa serwera wyszukiwania obrazów określona podczas instalacji.
6. W lewym okienku kliknij prawym przyciskiem myszy ikonę **Bazy danych**, a następnie wybierz opcję **Nowa baza danych**.
7. W oknie Nowa baza danych w polu **Nazwa** wpisz SAMPLEDB, a następnie kliknij przycisk **OK**.
8. W lewym okienku kliknij ikonę **Bazy danych**, aby wyświetlić ikonę SAMPLEDB.
9. Kliknij ikonę SAMPLEDB.
10. W lewym okienku kliknij prawym przyciskiem myszy ikonę **Katalogi**, a następnie kliknij pozycję **Nowy katalog**.
11. W oknie Nowy katalog w polu **Nazwa** wpisz SAMPLECAT, a następnie kliknij przycisk **OK**.
12. W okienku po lewej stronie kliknij ikonę **Katalogi**, aby wyświetlić ikonę SAMPLECAT.
13. Kliknij ikonę SAMPLECAT.
14. W okienku po lewej stronie kliknij prawym przyciskiem myszy ikonę **Funkcje**, a następnie kliknij pozycję **Nowe cechy**.
15. W oknie Nowe cechy w polu **Nazwa** zaznacz każdą z cech i kliknij przycisk **Zastosuj**.  
Po wybraniu wszystkich czterech cech zostanie wyświetlony następujący komunikat:  
Wszystkie możliwe cechy zostały dodane do katalogu.
16. Kliknij przycisk **OK**.
17. Kliknij przycisk **Anuluj**.

## Uruchamianie programu ładującego

Aby przetestować funkcje wyszukiwania tekstowego i wyszukiwania obrazów, należy załadować przykładowe dane.

Przykładowe dane graficzne znajdują się w następujących plikach:

**W systemie Windows:**

x:\cmbroot\samples\java\dl\samples.jar

**W systemie AIX:**

/usr/lpp/cmb/samples/java/dl/samples.jar

Przykładowy program ładujący załaduje dane do programu Content Manager, a następnie je zaindeksuje. Instrukcje dotyczące składni używanej do uruchamiania programu można znaleźć w początkowej części jego pliku źródłowego. Przykładowe programy ładujące to:

**W systemie Windows:**

x:\cmbroot\samples\java\dl\LoadSampleTSQBICDL.jar

**W systemie AIX:**

/usr/lpp/cmb/samples/java/dl/LoadSampleTSQBICDL.jar

Aby uruchomić program ładujący przykładowe dane:

1. Dokonaj dekompresji plików .jar, wprowadzając następującą komendę:  
jar -xvf samples.jar

Pliki zostaną rozpakowane do właściwych katalogów.

2. Ustaw zmienne środowiska stacji roboczej w sposób podany poniżej tak, aby skompilować przykładowy program ładujący:

**W systemie Windows:**

- a. Otwórz plik `x:\cmbroot\cmbenv71.bat` w edytorze tekstu i zmień pierwsze trzy wiersze, aby ustawić zmienne środowiska stacji roboczej:  

```
set CMBROOT = e:\cmbroot
set DB2HOME = e:\sql1lib
set JAVAHOME = d:\jdk117
```
- b. Zapisz plik `cmbenv71.bat` i ustaw zmienne środowiska, wprowadzając komendę:  
`cmbenv71`

**W systemie AIX:**

- a. Przejdź do katalogu `/usr/lpp/cmb/bin/` i uruchom program instalacyjny komendą  
`. /cmbenv71.sh`
  - b. Upewnij się, że wszyscy użytkownicy mają prawo zapisu do podkatalogów katalogu `/usr/lpp/cmb/samples/java/dl/` i przykładowych plików.
3. Skompiluj program ładujący, wprowadzając następującą komendę (wielkość liter jest istotna):  
`javac LoadSampleTSQBICDL.java`
  4. **Wymaganie:** Przed uruchomieniem programu ładującego trzeba uruchomić następujące serwery:
    - Serwer biblioteczny
    - Serwer obiektów
    - Serwer wyszukiwania tekstowego
    - Serwer wyszukiwania obrazów

Jeśli program Content Manager jest uruchomiony w wersji narodowej, przed uruchomieniem programu ładującego należy ustawić zmienną `FRNDEFLANG` na wartość `ENU`. W systemie AIX należy to wykonać przy użyciu komendy: `export FRNDEFLANG=ENU`

5. Załaduj przykładowe dane za pomocą programu ładującego, wpisując komendę:  
`java LoadSampleTSQBICDL sampleQBIC.dat load.log frnadmin password LIBSRVRN`

gdzie parametr `frnadmin` oznacza identyfikator użytkownika, `password` to hasło użytkownika, a `LIBSRVRN` to nazwa serwera bibliotecznego.

6. Sprawdź plik `load.log`, aby upewnić się, że przykładowe dane zostały załadowane prawidłowo.

Po zakończeniu ładowania przykładowych danych należy wykorzystać narzędzie administracyjne programu Content Manager lub funkcję wyszukiwania tekstowego uruchamianą z wiersza komend, aby zaindeksować przykładowe dane tekstowe.

## Indeksowanie przykładowych danych tekstowych

Aby zaindeksować przykładowe dane:

1. Uruchom program administrowania systemem Content Manager i zaloguj się do niego.
2. Z listy znajdującej się w lewym górnym okienku wybierz pozycję **Wyszukiwanie tekstowe**.
3. Kliknij dwukrotnie ikonę **Serwery wyszukiwania**.

4. Kliknij dwukrotnie ikonę **TM**. TM jest aliasem serwera wyszukiwania tekstowego.
5. Kliknij prawym przyciskiem myszy **Nowy indeks tekstowy**, a następnie kliknij opcję **Właściwości**.
6. W notatniku Właściwości na karcie Jawne kliknij przycisk **Odśwież**.
7. W polu **Liczba indeksów** powinna zostać wyświetlona liczba dokumentów, które zostały załadowane przy użyciu programu ładującego.
8. Kliknij przycisk **Indeks**, aby zaindeksować te pliki.
9. Oczekaj chwilę, a następnie kliknij przycisk **Odśwież**, aby w polu **Indeks podstawowych dokumentów** wyświetlić liczbę pomyślnie poindeksowanych dokumentów.

Po poindeksowaniu danych można tworzyć zapytania dotyczące kolekcji lub uruchamiać zapytania proste przy użyciu przykładowej aplikacji w języku Java lub narzędzia imlsrch uruchamianego z wiersza komend.





---

## Rozdział 9. Formaty dokumentów

---

### Formaty dokumentów programu eksploracji informacji

W tym załączniku opisano formaty dokumentów obsługiwane przez program pozyskiwanie informacji.

#### Przetwarzanie tekstu: Ogólne

<b>tekst ANSI (7 i 8 bitów)</b>	Wszystkie wersje
<b>tekst ASCII (dostępne wersje 7 i 8 bitowe)</b>	Wszystkie wersje
<b>HTML</b>	Wersje do 3.0 (z pewnymi ograniczeniami)
<b>IBM FFT</b>	Wszystkie wersje
<b>IBM Revisable Form Text</b>	Wszystkie wersje
<b>Microsoft Rich Text Format (RTF)</b>	Wszystkie wersje
<b>tekst Unicode</b>	Wszystkie wersje

#### Przetwarzanie tekstu: DOS

<b>DEC WPS Plus (DX)</b>	Wersje do 4.0
<b>DEC WPS Plus (WPL)</b>	Wersje do 4.1
<b>DisplayWrite 2 i 3 (TXT)</b>	Wszystkie wersje
<b>DisplayWrite 4 i 5</b>	Wersje do wydania 2.0
<b>Enable</b>	Wersje 3.0, 4.0 i 4.5
<b>First Choice</b>	Wersje do 3.0
<b>Framework</b>	Wersja 3.0
<b>IBM Writing Assistant</b>	Wersja 1.01
<b>Lotus Manuscript</b>	Wersje do 2.0
<b>MASS11</b>	Wersje do 8.0
<b>Microsoft Word</b>	Wersje do 6.0
<b>Microsoft Works</b>	Wersje do 2.0
<b>MultiMate</b>	Wersje do 4.0
<b>Navy DIF</b>	Wszystkie wersje
<b>Nota Bene</b>	Wersja 3.0
<b>Office Writer</b>	Wersje od 4.0 do 6.0
<b>PC-File Letter</b>	Wersje do 5.0
<b>PC-File+ Letter</b>	Wersje do 3.0
<b>PFS:Write</b>	Wersje A, B i C

<b>Professional Write</b>	Wersje do 2.1
<b>Q&amp;A</b>	Wersja 2.0
<b>Samna Word</b>	Wersje do 4.0
<b>SmartWare II</b>	Wersja 1.02
<b>Sprint</b>	Wersja 1.0
<b>Total Word</b>	Wersja 1.2
<b>Volkswriter 3 i 4</b>	Wersje do 1.0
<b>Wang PC (IWP)</b>	Wersje do 2.6
<b>WordMARC</b>	Wersje do Composer Plus
<b>WordPerfect</b>	Wersje do 6.1
<b>WordStar</b>	Wersje do 7.0
<b>WordStar 2000</b>	Wersje do 3.0
<b>XyWrite</b>	Wersje do III Plus

## **Przetwarzanie tekstu: Międzynarodowe**

<b>JustSystems Ichitaro</b>	Wersje 5.0, 6.0, 8.0, 9.0 i 10.0
-----------------------------	----------------------------------

## **Przetwarzanie tekstu: Windows**

<b>AMI/AMI Professional</b>	Wersje do 3.1
<b>Corel WordPerfect for Windows</b>	Wersje do 9.0
<b>JustWrite</b>	Wersje do 3.0
<b>Legacy</b>	Wersje do 1.1
<b>Lotus WordPro (Win32 / platforma Intel)</b>	SmartSuite 96, 97 i Millennium
<b>Lotus WordPro (platforma Unix - tylko tekst)</b>	SmartSuite 97 i Millennium
<b>Microsoft Windows Works</b>	Wersje do 4.0
<b>Microsoft Windows Write</b>	Wersje do 3.0
<b>Microsoft Word 97</b>	Word 97
<b>Microsoft Word 2000</b>	Word 2000
<b>Microsoft Word dla Windows</b>	Wersje do 7.0
<b>Microsoft WordPad</b>	Wszystkie wersje
<b>Novell Perfect Works</b>	Wersja 2.0
<b>Novell WordPerfect for Windows</b>	Wersje do 7.0
<b>Professional Write Plus</b>	Wersja 1.0
<b>Q&amp;A Write for Windows</b>	Wersja 3.0
<b>WordStar for Windows</b>	Wersja 1.0

## Przetwarzanie tekstu: Macintosh

Microsoft Word	Wersje od 4.0 do 6.0
Microsoft Word 98	Word 98
WordPerfect	Wersje od 1.02 do 3.0
Microsoft Works	Wersje do 2.0
MacWrite II	Wersja 1.1

## Arkusze kalkulacyjne

VP Planner 3D	Wersja 1.0
Enable	Wersje 3.0, 4.0 i 4.5
First Choice	Wersje do 3.0
Framework	Wersja 3.0
Lotus 1-2-3 (DOS i Windows)	Wersje do 5.0
Lotus 1-2-3 for SmartSuite	SmartSuite 97 i Millennium
Lotus 1-2-3 Charts (DOS i Windows)	Wersje do 5.0
Lotus 1-2-3 (OS/2)	Wersje do 2.0
Lotus 1-2-3 Charts (OS/2)	Wersje do 2.0
Lotus Symphony	Wersje 1.0, 1.1 i 2.0
Microsoft Excel 97	Excel 97
Microsoft Excel 2000	Excel 2000
Microsoft Excel Macintosh	Wersje od 3.0 do 4.0, 98
Microsoft Excel Windows	Wersje od 2.2 do 7.0
Microsoft Excel Charts	Wersje od 2.x do 7.0
Microsoft Multiplan	Wersja 4.0
Microsoft Windows Works	Wersje do 4.0
Microsoft Works (DOS)	Wersje do 2.0
Microsoft Works (Mac)	Wersje do 2.0
Mosaic Twin	Wersja 2.5
Novell Perfect Works	Wersja 2.0
QuattroPro for DOS	Wersje do 5.0
QuattroPro for Windows	Wersje do 9.0
PFS:Professional Plan	Wersja 1.0
SuperCalc 5	Wersja 4.0
SmartWare II	Wersja 1.02

## Bazy danych

SmartWare II	Wersja 1.02
--------------	-------------

<b>Access</b>	Wersje do 2.0
<b>dBase</b>	Wersje do 5.0
<b>DataEase</b>	Wersje 4.x
<b>dBXL</b>	Wersja 1.3
<b>Enable</b>	Wersje 3.0, 4.0 i 4.5
<b>First Choice</b>	Wersje do 3.0
<b>FoxBase</b>	Wersja 2.1
<b>Framework</b>	Wersja 3.0
<b>Microsoft Windows Works</b>	Wersje do 4.0
<b>Microsoft Works (DOS)</b>	Wersje do 2.0
<b>Microsoft Works (Mac)</b>	Wersje do 2.0
<b>Paradox (DOS)</b>	Wersje do 4.0
<b>Paradox (Windows)</b>	Wersje do 1.0
<b>Personal R:BASE</b>	Wersja 1.0
<b>R:BASE 5000</b>	Wersje do 3.1
<b>R:BASE System V</b>	Wersja 1.0
<b>Q &amp; A</b>	Wersje do 2.0
<b>Reflex</b>	Wersja 2.0

## Pliki graficzne

<b>PNG - Portable Network Graphics Internet Format</b>	Wersja 1.0
<b>Binary Group 3 Fax</b>	Wszystkie wersje
<b>BMP (RLE, ICO, CUR i os/2 DIB)</b>	Windows
<b>CDR (jeżeli wewnątrz osadzony jest obraz TIFF)</b>	Coral Draw 2.0 – 9.0
<b>CGM - Computer Graphics Metafile</b>	ANSI, CALS, NIST, wersja 3.0
<b>CMX - Corel Clip Art Format</b>	Wersje od 5 do 6
<b>DCX (wielostronicowy PCX)</b>	Microsoft Fax
<b>DRW - Micrografx Designer</b>	Wersja 3.1
<b>DRW - Micrografx Draw</b>	Wersje do 4.0
<b>DXF (binarne i ASCII) AutoCAD Drawing Interchange Format</b>	Wersje do 14
<b>EMF</b>	Windows Enhanced Metafile
<b>EPS Encapsulated PostScript</b>	Jeżeli wewnątrz osadzony jest obraz TIFF
<b>FMV - FrameMaker graphics</b>	Format rastrowy i wektorowy do wersji 5.0
<b>FPX - Kodak Flash Pix</b>	-

<b>GDF - IBM Graphics Data Format</b>	Wersja 1.0
<b>GEM - Graphics Environment Manager Metafile</b>	Format rastrowy i wektorowy
<b>GIF - Graphics Interchange Format</b>	Compuserve
<b>GP4 - Group 4 CALS Format</b>	Typ I i typ II
<b>HPGL - Hewlett Packard Graphics Language</b>	Wersja 2.0
<b>IMG - GEM Paint</b>	-
<b>JFIF (JPEG nie w formacie TIFF)</b>	Wszystkie wersje
<b>JPEG - Joint Photographic Experts Group format</b>	Wszystkie wersje
<b>MET - OS/2 PM Metafile</b>	Wersja 3.0
<b>PBM - Portable Bitmap</b>	-
<b>Kodak Photo CD</b>	Wersja 1.0
<b>PCD - mapa bitowa PCX</b>	PC Paintbrush
<b>Perfect Works (Draw)</b>	Wersja Novell 2.0
<b>PGM - Portable Graymap</b>	-
<b>PIC - Lotus 1-2-3 Picture File Format</b>	-
<b>PICT1 i PICT2 (Raster)</b>	Standard Macintosh
<b>PIF - IBM Picture Interchange Format</b>	Wersja 1.0
<b>PNTG</b>	MacPaint
<b>PPM - Portable Pixmap</b>	-
<b>Progressive JPEG</b>	-
<b>PSP - Paintshop Pro (tylko Win32)</b>	Wersje 5.0 i 5.0.1
<b>RND - AutoShade Rendering File Format</b>	Wersja 2.0
<b>SDW Ami Draw Snapshot (Lotus)</b>	Wszystkie wersje
<b>SRS - Sun Raster File Format</b>	-
<b>Targa</b>	Truevision
<b>TIFF</b>	Wersje do 6
<b>TIFF CCITT Group 3 i 4</b>	Fax Systems
<b>VISO (dla wersji 4 tylko podgląd strony) Visio 4, 5, 2000</b>	Visio 4, 5, 2000
<b>WMF</b>	Windows Metafile

<b>WordPerfect Graphics [WPG i WPG2]</b>	Wersje do 2.0
<b>XBM - X-Window Bitmap</b>	Zgodne z x10
<b>XPM - X-Window Pixmap</b>	Zgodne z x10
<b>XWD - X-Window Dump</b>	Zgodne z x10

## Profesjonalne formaty graficzne

<b>PSD - Adobe Photoshop File Format</b>	Wersja 4.0
<b>AI - Adobe Illustrator File Format</b>	Wersje do 7.0
<b>CDR - Corel Draw</b>	Wersje do 8.0
<b>DSF - Micrografx Designer</b>	Windows 95, wersja 6.0
<b>DWG - AutoCAD Native Drawing Format</b>	Wersje od 12 do 14
<b>IGES - Initial Graphics Exchange Specification</b>	Wersja 5.1
<b>PDF - Portable Document Format</b>	Acrobat wersja 2.1, 3.0, 4.0, w tym wersja japońska
<b>PS - Postscript</b>	Level 2

## Formaty prezentacyjne

<b>Microsoft PowerPoint for Macintosh</b>	Wersje 4.0, 98
<b>Corel Presentations</b>	Wersje 8.0 i 9.0
<b>Novell Presentations</b>	Wersje 3.0 i 7.0
<b>Harvard Graphics for DOS</b>	Wersje 2.x i 3.x
<b>Harvard Graphics</b>	Wersje Windows
<b>Freelance 96</b>	Freelance 96
<b>Freelance for Windows 95</b>	SmartSuite 97 i Millennium
<b>Freelance for Windows</b>	Wersje 1.0 i 2.0
<b>Freelance for OS/2</b>	Wersje do 2.0
<b>Microsoft PowerPoint for Windows</b>	Wersje do 7.0
<b>Microsoft PowerPoint 97</b>	PowerPoint 97
<b>Microsoft PowerPoint 2000</b>	PowerPoint 2000

## Formaty skompresowane i zakodowane

<b>ZIP PKWARE</b>	Wersje do 2.0g
<b>GZIP</b>	-
<b>LZA Self Extracting Compress</b>	-

<b>LZH Compress</b>	-
<b>Microsoft Binder</b>	Wersja 7.0, Binder 97
<b>MIME (tekst poczty elektronicznej)</b>	-
<b>UUEncode</b>	-
<b>UNIX Compress</b>	-
<b>UNIX tar</b>	-

## Inne

<b>vCard Electronic Business Card</b>	Wersja 2.1
<b>Pliki wykonywalne (EXE, DLL)</b>	-
<b>Pliki wykonywalne dla systemu Windows NT</b>	-
<b>MSG (tylko tekst)</b>	Format pocztowy programu Microsoft Outlook
<b>Microsoft Project (tylko tekst)</b>	Project 98





---

## Rozdział 10. Zarządzanie prawami

W tym rozdziale omówiono opcję zarządzania prawami portalu EIP oraz opisano koncepcje z tym związane. Wyjaśniono także techniki znakowania, których można użyć do ochrony własności niematerialnych i prawnych.

---

### Ochrona własności intelektualnej

Obiekty multimedialne przechowywane w formie elektronicznej stanowią własność intelektualną. Ochrona tych obiektów może okazać się newralgiczna dla działalności firmy, w szczególności w sytuacji, gdy są one prezentowane na stronach WWW, których kopiowanie jest względnie łatwe. Produkt Content Manager udostępnia technologię znakowania, której można użyć w celu uniemożliwienia nieuprawnionego korzystania z własności intelektualnej. Polega ona na znakowaniu cyfrowych obiektów multimedialnych lub umieszczaniu w nich odcisków umożliwiających identyfikację.

Cenne obiekty można znakować, aby:

- Możliwe było określenie źródła, co utrudni kopiowanie obiektu i jego wykorzystanie bez zezwolenia. Technika ta nosi nazwę znaku wodnego i jest zwykle widoczna dla użytkownika.
- Zidentyfikować odbiorcę zawartości, co utrudni kopiowanie i wykorzystanie bez zezwolenia. Technika ta nosi nazwę odcisku i jest zwykle niewidoczna.
- Umieścić w obiekcie dane autora, umożliwiając uzyskanie o nim dalszych informacji.
- Umieścić takie informacje, jak godzina i data, które można wykorzystać w łańcuchu dystrybucji.

Przed dostarczeniem obiektu cyfrowego klientom można go oznakować. Zarówno znak wodny, jak i odcisk można zastosować przed dostarczeniem. Jednakże zastosowanie odcisku przed dostarczeniem obiektu zakłada, że odbiorca jest znany, dlatego w niektórych przypadkach znakowanie musi być stosowane dynamicznie w procesie dostarczania. Zastosowanie znakowania w wewnętrznym środowisku firmy, przed dostarczeniem obiektu, zwiększa ochronę dzięki zmniejszeniu ryzyka nieuprawnionej manipulacji.

Znakowanie można stosować na wielu etapach procesu zarządzania i dostawy. Właściwe postępowanie zależy od konkretnej sytuacji. Znaczniki można stosować na następujących etapach przetwarzania:

- Przed zapisaniem obiektu

Jeśli obiekt ma zostać oznakowany w sposób typowy (na przykład widocznym znakiem wodnym, identyfikującym właściciela), można zastosować znakowanie przed zapisaniem lub w trakcie zapisywania obiektu. W systemie Content Manager można zapisać oba obiekty: oryginalny i oznakowany. Można także zapisać tylko oznakowany obiekt, a nieoznakowany umieścić w innym repozytorium.

- Po zapisaniu obiektu

Aby uniknąć znakowania obiektów znajdujących się w systemie Content Manager, można je wczytać, oznakować i zastąpić nieoznakowany obiekt wersją oznakowaną lub zapisać oznakowany obiekt jako nowy element.

- Podczas wydobywania obiektu

Jeśli sposób oznakowania zależy od odbiorcy, można czynność tę wykonać dynamicznie po wydobyciu obiektu. Wtedy zamiast oryginalnego obiektu zostanie dostarczony obiekt oznakowany.

Jeśli w systemie jest wiele nieoznakowanych obiektów pochodzących z poprzednich jego wersji, a użytkownik nie chce marnować czasu i zasobów na znakowanie każdego obiektu, czynność tę można wykonać podczas wydobywania obiektów.

---

## Korzystanie z technologii znakowania

Dostępnych jest wiele technologii znakowania. Każda z nich rozwiązuje inny problem i ma różny stopień odporności na usunięcie lub modyfikację.

Techniki znakowania można podzielić według:

- Rodzaju informacji

### **Znak wodny**

Identyfikuje źródło zawartości obiektu. Może zawierać takie informacje, jak właściciel i wersja obiektu.

**Odcisk** Identyfikuje odbiorcę zawartości obiektu. Może zawierać takie informacje, jak miejsce dostawy i odbiorcę.

- Widoczność

### **Widoczne**

Znakowanie jest widoczne.

### **Niewidoczne**

Znacznik jest ukryty w obrazie.

- Integralność

### **Delikatne**

Jakakolwiek modyfikacja powoduje uszkodzenie znacznika.

### **Odporne**

Znacznik jest odporny na takie modyfikacje obiektu, jak zmiana wielkości, kompresja, obrót i obcinanie.

- Czas zastosowania

- Podczas utrwalania obiektu
- Podczas zapisywania obiektu
- Podczas wydobywania obiektu w celu przesłania do klienta
- Podczas odbierania obiektu przez stację roboczą odbiorcy

- Położenie

- Jeśli widoczne oznakowanie ma przeciwdziałać nielegalnemu wykorzystaniu, można go zastosować do znacznej części obrazu. Można go przesunąć, aby pokrywał obszar obrazu o gęstszej fakturze, co utrudni jego usunięcie.
- Jeśli zostanie użyte oznakowanie niewidoczne, obszar o gęstej fakturze umożliwia umieszczenie danych w sposób mający nieznaczny wpływ na wygląd obrazu.
- Jeśli widoczne oznakowanie ma informować o prawach autorskich, można je umieścić w rogu obrazu, nie przesłaniając go.
- Jeśli ma być użyte zarówno oznakowanie widoczne, jak i niewidoczne, widoczne należy zastosować najpierw.

- Format

### **Binarne**

Znacznik może być przypadkowym łańcuchem bitów powtórzonym w pewnych odstępach. Łańcuch ten jest kluczem, za pomocą którego można zaznaczyć lub anulować zaznaczenie obrazu.

Znacznik może też być obrazem.

### **Dane strukturalne**

Znacznik może być osadzony danymi tekstowymi.

## **Oznakowanie widoczne**

Oznakowanie widoczne jest przezroczystą maską umieszczoną na obrazie w taki sposób, że widoczny jest zarówno obraz, jak i maska. Trudne do usunięcia widoczne oznakowanie skutecznie zapobiega przypisaniu autorstwa obiektów komu innemu.

Oznakowania widocznego używa się w następujących sytuacjach:

- Aby udostępnić klientom obrazy do obejrzenia, ale zniechęcić ich do dalszego użycia tych kopii.
- Aby użyć obrazu w reklamie w sieci WWW.

## **Oznakowanie niewidoczne**

Oznakowanie niewidoczne są to dane ukryte w obrazie w taki sposób, że wygląda on na niezmieniony. Aby umieścić znacznik, a następnie go wykryć i odcodować, potrzebna jest odpowiednia aplikacja.

Oznakowania niewidocznego używa się w następujących sytuacjach:

- Aby umieścić w obrazie informacje o prawach własności i utrudnić tworzenie nielegalnych kopii oznakowanych obiektów (znak wodny).
- Aby umieścić w obrazie informacje umożliwiające śledzenie ścieżki dystrybucji (odcisk).
- Aby umieścić w obrazie adnotację lub podpis.



---

## Rozdział 11. Funkcje dostępności

Produkt ten zawiera wiele funkcji, które ułatwiają używanie tego produktu przez osoby niepełnosprawne. Są to:

- Możliwość obsługi programu za pomocą klawiatury bez używania myszy.
- Obsługa rozszerzonych właściwości wyświetlania.
- Kompatybilność z technologiami asysty.
- Kompatybilność z funkcjami dostępności systemu operacyjnego.
- Przyjazne formaty dokumentacji.

---

### Wprowadzanie danych z klawiatury i nawigowanie

Do wprowadzania danych z klawiatury i nawigowania dostępne są następujące opcje:

#### Wprowadzanie danych z klawiatury

Produkt może być w pełni używany bez myszy.

Pozycje menu i elementy kontrolne są dostępne za pomocą klawiszy skrótów, dzięki czemu możliwe jest ich wybieranie i aktywacja bezpośrednio z klawiatury. Klawisze skrótów oznaczane są podkreśleniem w pozycji menu lub w elemencie kontrolnym.

#### Aktywowanie obiektów

W systemie Windows obiekt, który jest aktywny i może odbierać znaki wpisywane z klawiatury, jest wyróżniony.

#### Regulowanie czasu odpowiedzi

W systemie Windows można regulować czas odpowiedzi za pomocą Panelu sterowania.

---

### Funkcje wyświetlania

Programy klienckie są wyposażone w wiele funkcji rozszerzających interfejs użytkownika i poprawiających jego dostępność dla osób słabo widzących. Rozszerzenia te, to m.in. ustawienia wysokiego kontrastu i czcionki.

#### Tryb wysokiego kontrastu

Programy klienckie mogą używać trybu wysokiego kontrastu, który jest dostarczany przez system operacyjny. Polega on na zwiększeniu kontrastu między kolorami tła i pierwszego planu.

#### Ustawienia czcionek

W systemie Windows można określać ustawienia koloru, wielkości i kroju czcionki używanej w menu i oknach dialogowych. Program kliencki umożliwia wybranie czcionki z listy dokumentu.

#### Niezależność od koloru

Do używania tego produktu nie jest konieczne rozróżnianie kolorów przez obsługującą osobę.

---

### Kompatybilność z technologiami asysty

Programy klienckie są kompatybilne z programami czytającymi z ekranu, takimi jak Narrator i Via Voice. Programy klienckie mają właściwości wymagane przez tego typu aplikacje, umożliwiając odczytywanie informacji z ekranu użytkownikom słabo widzącym.

---

## Dostępna dokumentacja

Dokumentacja tego produktu dostarczana jest w postaci plików PDF. Pliki PDF można przekształcić do plików HTML za pomocą darmowych narzędzi, dostępnych na stronie producenta pod adresem: <http://access.adobe.com>. Dzięki temu możliwe jest przeglądanie dokumentacji zgodnie z ustawieniami określonymi w przeglądarce. Umożliwia to również działanie programów czytających z ekranu i innych technologii asysty.



---

## Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych.

IBM może nie oferować w innych krajach produktów, usług lub opcji omawianych w tej publikacji. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem, że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi, pochodzących od producenta innego niż IBM, spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na towary i usługi, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przysyłać na adres:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

Zapytania w sprawie licencji na informacje dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub zgłaszać na piśmie pod adresem:

IBM World Trade Asia Corporation  
Licensing  
2-31 Roppongi 3-chome, Minato-ku  
Tokyo 106, Japan

**Poniższy akapit nie obowiązuje w Wielkiej Brytanii, a także w innych krajach, w których jego treść pozostaje w sprzeczności z przepisami prawa miejscowego:**  
INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS"), BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (W TYM TAKŻE RĘKOJMI), WYRAŻNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA TA NIE NARUSZA PRAW OSÓB TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w niniejszej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych innych firm zostały wprowadzone wyłącznie dla wygody użytkownika i w żadnym wypadku nie stanowią zachęty

do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przysyłanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Informacje na temat możliwości stosowania tego programu, takie jak: (i) wymiana informacji między niezależnie tworzonymi programami a innymi programami (włącznie z tym programem) czy (ii) wspólne używanie wymienianych informacji, można uzyskać pod adresem:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003  
U.S.A.

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, uiszczenie odpowiedniej opłaty.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Wszelkie dane dotyczące wydajności zostały zebrane w kontrolowanym środowisku. W związku z tym rezultaty uzyskane w innych środowiskach operacyjnych mogą się znacząco różnić. Niektóre pomiary mogły być dokonywane na systemach będących w fazie rozwoju i nie ma gwarancji, że pomiary te wykonane na ogólnie dostępnych systemach dadzą takie same wyniki. Niektóre z pomiarów mogły być estymowane przez ekstrapolację. Rzeczywiste wyniki mogą być inne. Użytkownicy powinni we własnym zakresie sprawdzić odpowiednie dane dla ich środowiska.

Informacje dotyczące produktów firm innych niż IBM pochodzą od dostawców tych produktów, z opublikowanych przez nich zapowiedzi lub innych powszechnie dostępnych źródeł. Firma IBM nie testowała tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z tymi produktami. Pytania dotyczące produktów firm innych niż IBM należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennych operacjach działalności gospodarczej. W celu kompleksowego ich zilustrowania, podane przykłady zawierają nazwiska osób prywatnych, nazwy przedsiębiorstw oraz nazwy produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiegokolwiek podobieństwo do istniejących nazw/nazwisk i adresów jest całkowicie przypadkowe.

#### LICENCJA NA PRAWA AUTORSKIE:

Niniejsza publikacja zawiera przykładowe aplikacje w kodzie źródłowym, ilustrujące techniki programowania w różnych systemach operacyjnych. Użytkownik może kopiować, modyfikować i dystrybuować te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu projektowania, używania, sprzedaży lub dystrybucji aplikacji zgodnych z aplikacyjnym interfejsem programowym dla tego systemu operacyjnego, dla

którego napisane zostały programy przykładowe. Programy przykładowe nie zostały gruntownie przetestowane. IBM nie może zatem gwarantować lub sugerować niezawodności, użyteczności i funkcjonalności tych programów. Użytkownik może kopiować, modyfikować i dystrybuować te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu projektowania, używania, sprzedaży lub dystrybucji aplikacji zgodnych z aplikacyjnym interfejsem programowym IBM.

---

## Znaki towarowe

Następujące znaki są znakami towarowymi International Business Machines Corporation w Stanach Zjednoczonych i/lub w innych krajach:

IBM	DisplayWrite	PowerPC
400	e-business	PTX
Advanced Peer-to-Peer Networking	HotMedia	QBIC
AIX	Hummingbird	RS/6000
AIXwindows	ImagePlus	SecureWay
APPN	IMS	SP
AS/400	Micro Channel	VideoCharger
C Set ++	MQSeries	Visual Warehouse
CICS	MVS/ESA	VisualAge
DATABASE 2	NetView	VisualInfo
DataJoiner	OS/2	WebSphere
DB2	OS/390	
DB2 Universal Database	PAL	

Approach, Domino, Lotus, Lotus 1-2-3, Lotus Notes i SmartSuite są znakami towarowymi lub zastrzeżonymi znakami towarowymi Lotus Development Corporation w Stanach Zjednoczonych i/lub w innych krajach.

Intel i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows i Windows NT są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące języka Java są znakami towarowymi lub zastrzeżonymi znakami towarowymi Sun Microsystems, Inc. w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym The Open Group w Stanach Zjednoczonych i w innych krajach.

Nazwy innych firm, produktów i usług mogą być znakami towarowymi lub znakami usług innych podmiotów.



---

## Glosariusz

W tym glosariuszu zostały zdefiniowane terminy i skróty specyficzne dla opisywanego oprogramowania. Terminy oznaczone *kursywą* są zdefiniowane w glosariuszu w innym miejscu.

### A

**ADSM.** Patrz *Tivoli Storage Manager*.

**API.** Patrz *aplikacyjny interfejs programowy*.

**aplikacja kliencka.** Aplikacja napisana w języku obiektowym lub przy użyciu internetowego API, uzyskująca dostęp do *serwerów danych* z Enterprise Information Portal.

**aplikacyjny interfejs programowy (API).** Interfejs, który umożliwia wzajemną komunikację między aplikacjami. Interfejs API jest zestawem konstrukcji lub instrukcji języka programowania, które można umieścić w programie użytkowym w celu udostępnienia w nim określonych funkcji i usług, dostępnych w bazowym programie licencjonowanym.

**architektura IOCA (Image Object Content Architecture).** Zestaw konstruktorów wykorzystywanych do wymieniań i prezentacji obrazów.

**archiwizator multimediów .** Urządzenie fizyczne służące do przechowywania strumieni danych audio i wideo. Typem archiwizatora multimediów jest urządzenie VideoCharger.

**atrybut.** Jednostka danych, która opisuje określoną cechę lub właściwość (np. nazwę, adres, wiek, itp.) elementu i która może być używana do jego odszukania. Atrybut posiada typ, który oznacza zakres informacji określanych za pomocą atrybutu oraz wartość, która mieści się w zadanym zakresie. Na przykład w multimedialnym systemie plików informacje na temat pliku mogą zawierać tytuł utworu, czas trwania, rodzaj kodowania itp. Dla Enterprise Information Portal patrz także *atrybut stowarzyszony* i *atrybut rodzimy*.

**atrybut rodzimy.** Parametr obiektu zarządzanego przez określony *serwer danych*, który dotyczy tylko tego serwera danych. Na przykład *pole kluczowe nr polisy* może być atrybutem rodzimym na serwerze danych Content Manager, a *pole ID polisy* może być atrybutem rodzimym na serwerze danych Content Manager OnDemand.

**atrybut stowarzyszony.** Kategoria metadanych programu Enterprise Information Portal, która jest odwzorowywana na *atrybuty rodzime* na jednym lub kilku *serwerach danych*. Na przykład atrybut stowarzyszony numer polisy może zostać odwzorowany w programie Content Manager na *atrybut nr polisy*, a w programie Content Manager ImagePlus for OS/390 na atrybut ID polisy.

**Audio/Video Interleaved (Przepleciony obraz z dźwiękiem - AVI).** Specyfikacja pliku typu RIFF (*Resource Interchange File Format - Format pliku do wymiany zasobów*) umożliwiająca umieszczenie w jednym pliku przeplecionych danych dźwiękowych i obrazowych. Oddzielne ścieżki mogą być pobierane w kolejnych porcjach zarówno podczas odtwarzania, jak i nagrywania, dzięki czemu zachowany zostaje sekwencyjny dostęp do pliku.

**AVI.** Patrz *Audio/Video Interleaved*.

### B

**BLOB.** Patrz *duży obiekt binarny*.

**brama.** Jednostka funkcjonalna, która łączy dwie sieci komputerowe o różnej architekturze sieciowej. Most jest urządzeniem służącym do łączenia sieci lub systemów o takiej samej lub podobnej architekturze.

### C

**CGI.** Patrz *Common Gateway Interface*.

**CGI (Common Gateway Interface, standardowy interfejs bramki).** Standard wymiany informacji między serwerem WWW a programami zewnętrznymi. Programy zewnętrzne mogą być napisane w dowolnym języku programowania, który jest obsługiwany przez system operacyjny, pod kontrolą którego działa serwer WWW. Patrz *skrypt CGI*.

**CIF.** Patrz *standardowy plik wymiany danych*.

**CIU.** Patrz *standardowa jednostka wymiany danych*.

**część.** Patrz *obiekt*.

### D

**dane strumieniowe .** Jakikolwiek dane przesłane poprzez sieć z ustaloną szybkością. W strumieniu można przesłać dane jednego lub wielu typów. Szybkość transmisji wyrażona w bitach na sekundę może być różna dla różnych typów strumieni i sieci.

**DDO.** Patrz *dynamiczny obiekt danych*.

**definicja serwera.** Parametry określonego *serwera danych*, które jednoznacznie identyfikują go w programie Enterprise Information Portal.

**definicja typu dokumentu (DTD).** Reguły, które określają strukturę określonej klasy dokumentów XML. DTD definiuje strukturę za pomocą elementów, atrybutów i notacji oraz ustanawia ograniczenia stosowania elementów, atrybutów i

notacji w dokumencie. DTD można porównać do schematu bazy danych; DTD całkowicie opisuje strukturę danego języka znaczników.

**definicja typu serwera.** Określona przez administratora lista parametrów wymaganych do jednoznacznej identyfikacji w programie Enterprise Information Portal niestandardowego serwera określonego typu.

**dokument.** *Element*, który można przechowywać, pobierać i wymieniać między programami Content Manager i użytkownikami jako pojedynczą jednostkę. Element z dokumentem należącym do *typu semantycznego* powinien zawierać informacje tworzące dokument, ale niekoniecznie musi to być implementacja modelu dokumentu systemu Content Manager.

Element utworzony z typu elementu sklasyfikowanego (odmiana implementacji modelu dokumentu Content Manager) musi zawierać części dokumentu. Tego rodzaju dokumenty mogą być używane do tworzenia elementów z typem semantycznym dokumentu lub folderu.

Części dokumentu mogą zawierać różne typy danych, na przykład tekst, obrazy i arkusze kalkulacyjne.

**DTD.** Patrz *definicja typu dokumentu*.

**duży obiekt binarny (BLOB).** Sekwencja bajtów o rozmiarach od 0 bajtów do 2 gigabajtów. Łańcuch znaków, z którym nie jest powiązana strona kodowa ani zestaw znaków. Obiekty obrazów, audio i wideo są przechowywane w obiektach BLOB.

**dynamiczny obiekt danych (dynamic data object, DDO).** Ogólna reprezentacja programowa przechowywanego obiektu służąca do przenoszenia tego obiektu do i z miejsca składowania.

## E

**eksploracja informacji.** Zautomatyzowany proces wydobywania z tekstu kluczowych informacji (podsumowywanie), znajdowania dominujących motywów w kolekcji dokumentów (klasyfikacja) i wyszukiwania odpowiednich dokumentów przy użyciu efektywnych i elastycznych zapytań.

**element.** Ogólne określenie najmniejszej jednostki informacji, którą zarządza program Enterprise Information Portal. Każdy element ma identyfikator. Na przykład typem elementu może być *folder* lub *dokument*.

**element pracy.** We wcześniejszych wersjach programu Content Manager i Enterprise Information Portal czynność aktywna w *przepływie pracy*.

**etap.** Punkt w *przepływie pracy* lub *przepływie dokumentów*, przez który musi przejść każdy *element pracy*, *dokument* lub *folder*.

**Extensible Markup Language (XML).** Standard metajęzyka służącego do definiowania języków znacznikowych, który

wywodzi się (i jest podzbiorem) języka SGML. W języku XML pominięto bardziej złożone i rzadko używane części języka SGML, dzięki czemu jest on łatwy do stosowania podczas tworzenia aplikacji obsługujących typy dokumentów, zarządzających informacją hierarchiczną oraz przesyłających i współużytkujących tę informację w różnych systemach komputerowych. Użycie języka XML nie wymaga stabilnych aplikacji i przetwarzania, które są niezbędne w przypadku języka SGML. Język XML jest rozwijany pod patronatem W3C (World Wide Web Consortium).

## F

**folder.** *Element* należący do dowolnego *typu elementu* niezależnie od klasyfikacji, należący do *typu semantycznego* folderu. Każdy folder, który ma typ semantyczny folderu obsługiwany przez produkt Content Manager, ma możliwości obsługi elementów nie będących zasobami oraz inne możliwości dostępne dla klasyfikacji elementu, który może być *dokumentem* albo elementem zasobu. Foldery mogą zawierać dowolną liczbę elementów każdego typu, w tym dokumenty i podfoldery. Folder może być indeksowany przy użyciu *atributów*.

**format danych.** Patrz *typ MIME*.

**format pliku do wymiany zasobów (Resource Interchange File Format - RIFF).** Format pliku używany do przechowywania danych dźwiękowych lub graficznych w celu odtwarzania na różnego typu komputerach.

**funkcja.** Informacje wizualne, przechowywane na serwerze wyszukiwania obrazów. Także cecha wizualna wykorzystywana przez aplikacje wyszukiwania obrazów do sprawdzania zgodności ze wzorcem. Cztery cechy, których może dotyczyć zapytanie *QBIC* to: przeciętny kolor, kolor histogramu, kolor pozycyjny i tekstura.

## G

**GB.** Patrz *gigabajt*.

**gigabajt (GB).** (1) W pamięci rzeczywistej lub wirtualnej komputera: 2<sup>30</sup> lub 1 073 741 824 bajtów. (2) W urządzeniu pamięci masowej i w komunikacji: 1 000 000 000 bajtów.

**grupa użytkowników.** Grupa składająca się z jednego lub wielu indywidualnych *użytkowników*, identyfikowana za pomocą jednej nazwy grupy.

## H

**HTML.** Patrz *Hypertext Markup Language*.

## I

**identyfikator trwały (persistent identifier, PID).**

Identyfikator, który jednoznacznie określa *obiekt*, bez względu na miejsce jego przechowywania. Identyfikator PID składa się z identyfikatora elementu i położenia.



**indeksowanie.** Dodawanie lub edytowanie wartości atrybutu, który identyfikuje określone *elementy* lub *obiekty* w celu późniejszego pobrania.

**IOCA.** Patrz *architektura IOCA*

**iterator.** Klasa lub konstrukcja używana do przechodzenia przez kolejne obiekty kolekcji.

## J

**JavaBeans.** Niezależna od platformy technika programowania służąca do budowania nadających się do wielokrotnego wykorzystywania komponentów języka Java o nazwie “beans”. Składniki bean można udostępniać innym programistom lub używać w aplikacjach pisanych przy użyciu języka Java. Składniki JavaBeans mogą być również przekształcane i osadzone w graficznym środowisku programowania umożliwiającym przeciąganie i upuszczanie.

**jednostka rodzima.** *Obiekt zarządzany przez określony serwer danych i składający się z atrybutów rodzimych.* Na przykład klasy indeksowe programu Content Manager są jednostkami rodzimymi składającymi się z *pól kluczowych* programu Content Manager.

**jednostka stowarzyszona.** Obiekt metadanych programu Enterprise Information Portal składający się z *atrybutów stowarzyszonych* i opcjonalnie powiązany z jednym lub kilkoma *stowarzyszonymi indeksami tekstowymi*.

**język HTML (Hypertext Markup Language).** Język znaczników zgodny ze standardem SGML, przeznaczony głównie do obsługi wyświetlania w trybie online informacji tekstowych i graficznych, zawierających odsyłacze hipertekstowe.

**Joint Photographic Experts Group (JPEG).** (1) Grupa ustanawiająca standardy kompresji obrazów. (2) Standard zapisywania obrazów w pliku.

**JPEG.** Patrz *Joint Photographic Experts Group*.

## K

**klasa.** Model lub szablon w programowaniu lub projektowaniu obiektowym, który można wywołać w celu utworzenia obiektów mających wspólną definicję, a w związku z tym wspólne właściwości, operacje i działanie. Obiekt jest instancją klasy.

**klasa abstrakcyjna.** W programowaniu obiektowym jest to *klasa* reprezentująca pewną ideę, którą implementują klasy po niej dziedziczące. Nie można utworzyć obiektu klasy abstrakcyjnej; nie może ona posiadać własnych instancji.

**klasa indeksowa.** Patrz *typ elementu*.

**klasa połączeniowa.** Występująca w programowaniu obiektowym *klasa* umożliwiająca dostęp do API związanego z konkretnym *serwerem danych*.

**klasyfikacja typu elementu.** Podział na kategorie wewnątrz *typów elementu*, który służy do identyfikowania *elementów* tego typu. Wszystkie elementy tego samego typu mają taką samą klasyfikację.

Z programem Content Manager dostarczane są następujące klasyfikacje: *folder*, *dokument*, obiekt, film, obrazek i tekst, użytkownicy mogą również definiować własne klasyfikacje.

**klient biblioteczny.** Składnik programu Content Manager zapewniający niskopoziomowy interfejs programistyczny dla systemu bibliotecznego. Klient biblioteczny zawiera interfejsy API stanowiące część pakietu do projektowania oprogramowania.

**klient uszczuplony.** Klient bez oprogramowania lub z zainstalowanym minimalnym zestawem oprogramowania, który ma dostęp do oprogramowania zarządzanego i dostarczanego przez serwery sieciowe, z którymi jest połączony. Klient uszczuplony jest rozwiązaniem alternatywnym w stosunku do klienta o pełnych możliwościach, np. stacji roboczej.

**klient/serwer.** W komunikacji jest to model współdziałania w rozproszonym przetwarzaniu danych, gdzie program na jednym hoście wysyła żądanie do programu na innym hoście i oczekuje na odpowiedź. Program wysyłający żądanie jest nazywany klientem, a program odpowiadający - serwerem.

**kolekcja.** Grupa obiektów o zbliżonych zestawach reguł zarządzania.

**kolekcja stowarzyszona.** Grupowanie obiektów będące wynikiem *wyszukiwania stowarzyszonego*.

**komponent główny.** Pierwszy lub jedyny poziom w hierarchicznym *typie elementu*, zawierający *atrybuty* zdefiniowane przez użytkownika i systemowe.

**komponent potomny.** Opcjonalny drugi lub niższy poziom w hierarchii *typów elementów*. Każdy komponent potomny jest bezpośrednio skojarzony z poziomem nadrzędnym.

**konstruktor.** Metoda dotycząca języków programowania o tej samej nazwie co klasa, która służy do tworzenia i inicjowania obiektów danej klasy.

**kontrola dostępu .** Proces, który zapewnia, że dostęp do określonych funkcji i przechowywanych *obiektów* mają wyłącznie autoryzowani użytkownicy w sposób dozwolony w systemie.

**kryteria wyszukiwania.** W programie Enterprise Information Portal - określone pola zdefiniowane przez administratora dla *szablonu wyszukiwania*, które ograniczają lub dokładniej definiują możliwości dokonania wyboru przez *użytkownika*.

**kursor.** Nazwana struktura sterowania służąca aplikacji do wskazywania określonego wiersza w uporządkowanym zestawie wierszy. Kursor służy do pobierania wierszy z zestawu.



## L

**LAN.** Patrz *lokalna sieć komputerowa*.

**liczność.** Liczba wierszy w tabeli bazy danych.

**lista czynności.** Zatwierdzona lista czynności, zdefiniowana przez administratora systemu lub *koordynatora przepływu pracy*, określająca dozwolone czynności, które może wykonać użytkownik w *przepływie pracy* lub procesie przepływu dokumentów.

**lista kontroli dostępu.** Lista zawierająca jeden lub wiele identyfikatorów użytkowników lub grup użytkowników i przyznane im *uprawnienia*. Za pomocą list kontroli dostępu można również zarządzać dostępem użytkowników do *szyblonu wyszukiwania* w systemie Enterprise Information Portal.

**lista pracy.** Kolekcja *elementów pracy*, *dokumentów* lub *folderów*, które są przypisane użytkownikowi.

**lokalna sieć komputerowa (LAN).** Sieć połączonych ze sobą urządzeń, która może mieć połączenie z większą siecią.

## Ł

**łańcuch zapytania.** Łańcuch znaków określający właściwości i ich wartości w zapytaniu. Łańcuch zapytania można utworzyć w aplikacji i przekazać do odpowiedniego zapytania.

**łącze.** Kierunkowa relacja między dwoma *elementami*: źródłowym i docelowym. Za pomocą zbioru łączy można modelować powiązania typu jeden-wiele. Patrz termin przeciwstawny *odniesienie*.

**łączenie i osadzanie obiektów (Object Linking and Embedding, OLE).** Specyfikacja firmy Microsoft umożliwiająca łączenie i osadzanie aplikacji w taki sposób, aby możliwe było ich uaktywnianie z poziomu innych aplikacji.

## M

**menedżer folderów.** Model programu Content Manager do zarządzania danymi jako folderami i dokumentami elektronicznymi. Jako głównego interfejsu między aplikacjami a serwerami danych Content Manager można użyć interfejsu API menedżera folderów.

**menedżer zasobów.** Komponent programu Content Manager, który zarządza *obiektami*. Do obiektów tych odnoszą się *elementy* zapisane na *serwerze bibliotecznym*.

**metoda.** W języku Java kod implementujący funkcjonalność dostarczaną przez daną operację. Synonim funkcji obiektu w języku C++.

**multimedia.** Połączenie różnych przekazów (tekstu, grafiki, dźwięku, zdjęć, filmu i animacji) w celu wyświetlenia na komputerze.

**multimedialny system plików.** *System plików*, który został zoptymalizowany do przechowywania filmów i dźwięków.

**Multipurpose Internet Mail Extensions (MIME) .** Patrz *typ MIME*.

## N

**nadklasa.** *Klasa*, od której pochodzi dana klasa. Pomiedzy klasą a nadklasą mogą występować inne klasy.

**nakładka.** Zbiór zdefiniowanych danych, takich jak linie, cieniowanie, tekst, prostokąty lub logo, które podczas drukowania mogą być umieszczane na stronie z danymi zmiennymi.

## O

**obiekt.** Dowolna cyfrowa zawartość, którą użytkownik może zapisywać, odczytywać i manipulować jako jedną jednostkę, np. obrazy *JPEG*, dźwięk *MP3*, filmy *AVI* i tekst książki.

**obszar etapowy.** Roboczy obszar pamięci *menedżera zasobów*. Nazywany również *pamięcią podręczną menedżera zasobów*.

**odniesienie.** Powiązanie typu jeden-z-jednym między głównym komponentem lub *komponentem potomnym* i innym *głównym komponentem*. Patrz termin przeciwstawny *łącze*.

**odwzorowywanie użytkownika.** Wiązanie identyfikatorów użytkowników i haseł programu Enterprise Information Portal z odpowiadającymi im identyfikatorami użytkowników i hasłami jednego lub kilku serwerów danych. Odwzorowanie użytkownika umożliwia logowanie się w programie Enterprise Information Portal i na wielu *serwerach danych* przy użyciu jednej operacji logowania.

**OLE.** Patrz *łączenie i osadzanie obiektów*.

## P

**pakiet.** Kolekcja pokrewnych *klas* i interfejsów, które udostępniają ochronę dostępu i zarządzanie przestrzenią nazw.

**pakiet prac.** W programie Enterprise Information Portal w wersji 7.1 zbiór *dokumentów* przenoszony z jednego miejsca w inne. Użytkownicy używają pakietów prac za pośrednictwem *list prac*.

**pamięć podręczna.** Pamięć specjalnego przeznaczenia, mniejsza i szybsza od głównej pamięci, używana do przechowywania kopii danych, które są najczęściej przetwarzane. Użycie pamięci podręcznej skraca czas dostępu, może jednak zwiększać wymagania pamięciowe.

**pamięć podręczna >menedżera zasobów.** Roboczy obszar pamięci *menedżera zasobów*. Zwany również *obszarem etapowym*.

**pamięć podręczna serwera obiektów.** Patrz *pamięć podręczna menedżer zasobów*.

**PID.** Patrz *identyfikator trwały*.

**plik README.** Plik, który powinien być przeczytany przed uruchomieniem lub zainstalowaniem programu, którego dotyczy. Plik README zwykle zawiera ostatnie informacje na temat produktu, informacje dotyczące instalacji lub wskazówki używania produktu.

**plik tabeli sieciowej.** Plik tekstowy zawierający systemowe informacje konfiguracyjne dla każdego węzła programu Content Manager. W każdym węźle systemu musi znajdować się plik tabeli sieciowej identyfikujący ten węzeł i zawierający listę węzłów, z którymi trzeba nawiązywać połączenia.

Plik ten nosi nazwę FRNOLINT.TBL.

**podklasa.** Klasa pochodna innej klasy. Pomiedzy klasą a podklasą mogą występować inne klasy.

**podzbiór klasy indeksowej.** We wcześniejszych wersjach programu Content Manager, widok *klasy indeksowej*, który jest używany przez aplikację do zapisywania, pobierania oraz wyświetlania folderów i obiektów.

**pole kluczowe.** Patrz *atrybut*.

**procedura użytkownika.** Procedura napisana przez użytkownika, która przejmuje kontrolę we wcześniej zdefiniowanych punktach.

**program użytkownika.** Punkt w programie dostarczonem przez firmę IBM, w którym to punkcie sterowanie może zostać przejęte przez procedurę wyjściową użytkownika.

**protokół historii.** Plik, w którym są zapisywane czynności związane z *przepływem pracy*.

**przenoszenie etapowe.** Proces przenoszenia przechowywanego *obiektu* z urządzenia o niskim priorytecie (lub dostępnego offline) do urządzenia o wyższym priorytecie (lub dostępnego w trybie online), następujący zwykle na żądanie systemu lub użytkownika. Gdy użytkownik zażąda obiektu przechowywanego w pamięci trwałej, kopia robocza tego obiektu jest zapisywana w *obszarze etapowym*.

**przepływ pracy.** W programie Enterprise Information Portal sekwencja *czynności* i reguł zarządzających tymi czynnościami, przez które przechodzi *pakiet roboczy*, *dokument* lub *folder* podczas przetwarzania.

Na przykład zatwierdzenie wypłaty odszkodowania można opisać jako proces, w którym określone żądanie wypłaty musi nadejść w celu zatwierdzenia.

## Q

**QBIC.** Patrz *zapytanie o zawartość obrazu*.

## R

**ranga.** Liczba całkowita określająca stopień dopasowania danej części do wyników zapytania. Wyższa ranga oznacza większą zgodność.

**renderowanie.** Pobieranie danych, które nie są danymi graficznymi i wyświetlanie ich na ekranie w formie obrazu. W programie Content Manager można w ten sposób wyświetlać dokumenty utworzone w edytorach tekstowych.

**RIFF.** Patrz *format pliku do wymiany zasobów*.

**rodzimy indeks tekstowy.** Indeks *elementów* tekstu, który jest zarządzany przez określony *serwer danych*. Na przykład pojedynczy indeks wyszukiwania tekstowego na serwerze danych Content Manager.

**rozszerzony obiekt danych (extended data object, XDO).** Ogólna reprezentacja programowa przechowywanego złożonego *obiektu* multimedialnego służąca do przenoszenia tego obiektu do i z miejsca składowania. XDO są najczęściej częścią *DDO*.

## S

**serwer biblioteczny.** Komponent programu Content Manager, w którym są zapisywane, zarządzane i obsługiwane żądania dotyczące *elementów*.

**serwer danych.** System oprogramowania, w którym są przechowywane dane multimedialne i biznesowe oraz pokrewne metadane potrzebne użytkownikom podczas pracy z tymi danymi. Przykładami serwerów danych mogą być programy Content Manager i Content Manager ImagePlus for OS/390.

**serwer multimediiów.** Składnik programu Content Manager pracujący pod kontrolą systemu AIX, wykorzystywany do przechowywania i udostępniania plików wideo.

**serwer obiektów.** Patrz *zarządca zasobów*.

**serwer RMI.** Serwer, na którym jest zaimplementowany model obiektów rozproszonych *zdalnego wywoływania metod języka Java (RMI)*.

**składnica danych.** (1) Ogólny termin określający miejsce przechowywania danych (na przykład system baz danych, plik lub katalog). (2) W programie, wirtualna reprezentacja *serwera danych*.

**składnik.** Termin ogólny określający *składnik główny* lub *składnik potomny*.

**skrypt CGI.** Program komputerowy działający na serwerze sieci WWW, używający *interfejsu CGI (Common Gateway Interface, standardowy interfejs bramki)* do wykonywania zadań, które nie są zwykle wykonywane przez serwer WWW (na przykład dostęp do baz danych i przetwarzanie formularzy). Skrypt CGI jest programem CGI, który został napisany w języku skryptowym, np. w języku Perl.

**spis zasobów serwerów.** Obszerna lista *elementów rodzimych* i *atributów rodzimych* z określonych *serwerów danych*.

**stan pracy.** Status pojedynczego *elementu pracy, dokumentu* lub *folderu*.

**stan przepływu pracy.** Stan całego *przepływu pracy*.

**stowarzyszona składnica danych.** Wirtualna reprezentacja dowolnej liczby określonych *serwerów danych*, takich jak program Content Manager.

**stowarzyszony indeks tekstowy.** Obiekt metadanych programu Enterprise Information Portal, który jest odwzorowany na jeden lub kilka *rodzimych indeksów tekstowych* na jednym lub kilku *serwerach danych*.

**system plików.** W systemie AIX sposób podziału dysku twardego umożliwiający przechowywanie na nim danych.

**szablon wyszukiwania.** Zaprojektowany przez administratora formularz składający się z *kryteriów wyszukiwania*, służący do przeprowadzania określonego typu wyszukiwania stowarzyszonego. Administrator również nadaje *użytkownikom i grupom użytkowników* uprawnienia dostępu do każdego z szablonów wyszukiwania.

## T

**Tivoli Storage Manager (TSM).** Produkt typu *klient/serwer*, zapewniający usługi związane z zarządzaniem pamięcią i dostępem do danych w środowisku heterogenicznym. Obsługuje różnorodne metody komunikacji, udostępnia narzędzia administracyjne do zarządzania składowaniem i przechowywaniem plików oraz funkcje umożliwiające tworzenie harmonogramów operacji składowania.

**TSM.** Patrz *Tivoli Storage Manager*.

**typ elementu.** Szablon służący do definiowania i późniejszego wyszukiwania *elementów* składających się z *głównego komponentu*, zero lub kilka *komponentów potomnych* i klasyfikacji.

**typ MIME.** Internetowy standard służący do identyfikowania typów obiektów przesyłanych siecią Internet. Typy MIME definiują m.in. kilka różnych odmian dźwięku, obrazu i filmu. Każdy obiekt ma swój typ MIME.

**typ semantyczny.** Definiuje użycie lub reguły dla *elementu*. Podstawa, adnotacja i informacja to typy semantyczne dostarczane wraz z programem Content Manager; użytkownicy mogą też definiować własne typy.

## U

**uchwyt.** Łańcuch znaków reprezentujący obiekt i wykorzystywany do jego pobierania.

**uprawnienie.** Prawo do uzyskania dostępu do określonego *obiektu* w określony sposób. Uprawnienia obejmują prawa do

tworzenia, usuwania i wybierania obiektów przechowywanych w systemie. Uprawnienia są przyznawane przez administratora.

**URL.** Ciąg znaków reprezentujący zasoby informacyjne w komputerze lub sieci. Ciąg ten zawiera skrót nazwy protokołu, używanego do uzyskania dostępu do zasobu oraz informacje używane przez ten protokół do odszukania tego zasobu. Na przykład w internecie używane są m.in. następujące skróty protokołów: http, ftp, gopher, telnet i news.

**uwolnienie.** Usunięcie kryteriów powodujących zawieszenie *elementu*. Zawieszony element jest uwalniany po spełnieniu określonych kryteriów lub gdy użytkownik mający odpowiednie uprawnienia pominie te kryteria i dokona ręcznego uwolnienia elementu.

**użytkownik.** W przypadku programu Enterprise Information Portal jest to dowolna osoba zidentyfikowana w programie administracyjnym Enterprise Information Portal.

## V

**video-on-demand (VOD).** Usługa dostarczania konsumentom filmów i innych programów na żądanie.

## W

**widok klasy indeksowej.** We wcześniejszych wersjach programu Content Manager termin używany w interfejsach API na określenie *podzbioru klasy indeksowej*.

**właściwość.** Parametr *obiektu* opisujący ten obiekt. Właściwość można zmieniać lub modyfikować. Przykładem właściwości jest styl wpisywania.

**wolumin.** Reprezentacja rzeczywistego fizycznego urządzenia lub jednostki pamięci masowej, na którym są przechowywane obiekty w systemie.

**wolumin TSM.** Logiczny obszar pamięci zarządzany przez program *Tivoli Storage Manager*.

**wspólna jednostka wymiany (CIU).** Niezależna jednostka transferu dla wspólnego pliku wymiany (CIF, Common Interchange File). Jest to część pliku CIF określająca jego związek z docelową bazą danych. Plik CIF może zawierać wiele jednostek CIU.

**wspólny plik wymiany (CIF).** Plik zawierający jeden strumień danych IPIA (ImagePlus Interchange Architecture).

**wymiana.** Możliwość importowania lub eksportowania obrazu razem z jego indeksem z jednego systemu Content Manager ImagePlus for OS/390 do innego systemu ImagePlus za pośrednictwem *standardowego pliku wymiany danych* lub *standardowej jednostki wymiany danych*.

**wyszukiwanie parametryczne.** Zapytanie o *obiekty* oparte na *właściwościach* tych obiektów.

**wyszukiwanie stowarzyszone.** Zapytanie wystosowane z programu Enterprise Information Portal realizujące jednocześnie wyszukiwanie danych na kilku *serwerach danych*, które mogą być heterogeniczne.

**wyszukiwanie złożone.** Zapytanie łączące wyszukiwania jednego typu lub kilku typów spośród następujących: *parametryczne*, tekstowe lub wyszukiwanie obrazów.

## X

**XDO.** Patrz *rozszerzony obiekt danych*.

**XML.** Patrz *Extensible Markup Language*.

## Z

**zapytanie o zawartość obrazu (QBIC).** Technika zapytań umożliwiająca przeprowadzanie wyszukiwań w oparciu o zawartość wizualną, a nie zwykły tekst. Korzystając z zapytania QBIC, można wyszukiwać obiekty na podstawie ich wizualnych cech, takich jak kolor i tekstura.

**zawieszanie.** Usunięcie *obiektu* z jego *przepływu pracy* i określenie kryteriów zawieszenia wymaganych do jego uaktywnienia. Późniejsze uaktywnienie obiektu umożliwia kontynuowanie przetwarzania.

**zdalne wywoływanie metod (Remote Method Invocation, RMI).** Zestaw interfejsów API umożliwiający programowanie rozproszone. Obiektowi w jednej wirtualnej maszynie języka Java (JVM) umożliwia wywoływanie metod na obiektach w innych maszynach JVM.

**zestaw uprawnień.** Zbiór *uprawnień* do pracy z komponentami i funkcjami systemu. Zestawy uprawnień są przydzielane użytkownikom (identyfikatorom użytkowników) i *grupom użytkowników* przez administratora.

**znak zastępczy.** Znak specjalny, np. gwiazdka (\*) lub znak zapytania (?), który może być używany do reprezentowania jednego lub kilku znaków. Znak zastępczy może być zastępowany przez dowolny znak lub zbiór znaków.

## Ż

**żądanie.** Część adresu WWW, który występuje po *protokole* i *nazwie hosta*. Na przykład w *adresie* <http://www.serwer.com.pl/jz/foto.html>, żądaniem jest [/jz/foto.html](http://www.serwer.com.pl/jz/foto.html).



# Indeks

## C

cmbcc2mime.ini 16  
czynności, definiowanie 102

## D

domena 40  
domena administracyjna 36  
domeny  
    objaśnienie 37  
    tworzenie 37  
    uprawnienia podadministratora 38  
    uprawnienia superadministratora 38  
dostępność 133  
dostosowywanie typów MIME 16

## E

EIP  
    klient przeglądarki danych 6  
    klient wyszukiwania obrazów 5  
    klient wyszukiwania tekstowego 5  
    łączniki 5  
    opcja eksploracji informacji 5  
    opcja Web Crawler 5  
    pakiet narzędzi do programowania  
        łączników 6  
    składnik administracyjny 4  
    składnik Centrum informacyjne 7  
eksploracja informacji  
    grupa docelowa 46  
    obsługiwane formaty dokumentów 51  
    obsługiwane języki 51  
    praca w środowisku biznesowym 46  
    przykład 47  
    tworzenie systematyki 63  
    usługa 43  
Eksploracja informacji  
    opis 43  
    składniki 44  
Enterprise Information Portal  
    definiowanie  
        czynności 102  
        lista czynności 102  
        listy prac 101  
    tworzenie  
        kryteria wyszukiwania 28  
        przepływ pracy 102  
        szablony wyszukiwania 27

## G

grupa uprawnień 35  
grupa użytkowników 36  
    przenoszenie domen 40

## I

IBM Enterprise Information Portal for  
    Multitplatforms  
        składniki 3  
identyfikator użytkownika 31  
Information Structuring Tool  
    definiowanie systematyki 64  
    instalowanie 63  
    mechanizm blokowania 63  
    opis 63  
    pierwsze kroki 63  
    szkolenie systematyki 72  
    używanie WAS 63  
    wartościowanie systematyki 68  
    wybieranie dokumentów  
        szkoleniowych 66

## K

katalog  
    dodawanie 64  
    dodawanie dokumentów  
        szkoleniowych 66  
    szkolenie 72  
    usuwanie 64  
    wartościowanie 68  
    zmiana nazwy 64  
klawiatura 133  
klient administracyjny  
    definiowanie  
        czynności 102  
        lista czynności 102  
        listy prac 101  
    tworzenie  
        kryteria wyszukiwania 28  
        przepływ pracy 102  
        szablony wyszukiwania 27  
kolekcja  
    przypisywanie do domeny 39  
kolekcje  
    przenoszenie domen 41  
konfigurowanie wyszukiwania obrazów 112  
kryteria wyszukiwania  
    definiowanie i przypisywanie 28

## L

LDAP  
    importowanie 34  
    konfigurowanie 34  
lista czynności  
    definiowanie 99, 102  
    zdefiniowane czynności 102  
lista kontroli dostępu  
    przenoszenie domen 41  
lista prac  
    definiowanie 101  
    opis 99

## Ł

ładowanie dokumentów wyszukiwania  
    tekstowego i wyszukiwania obrazów 117  
ładowanie przykładowych danych 115  
łączniki 5

## M

menedżer zasobów  
    przypisywanie do domeny 39  
    przypisywanie użytkowników 36  
menedżer zasobów, przenoszenie do innej  
    domeny 40  
metadane  
    korzystanie z eksploracji informacji 43

## N

niedostępność 133

## O

opcja przeglądarki danych 6  
opcja wyszukiwania obrazów 5

## P

pakiet pracy, opis 98  
planowanie  
    Enterprise Information Portal 2  
plik typów MIME  
    zmienianie dla serwerów 16  
program budujący przepływ pracy  
    opis 100  
    tworzenie przepływu pracy 102  
uruchamianie 102  
przepływ pracy  
    konfigurowanie 95  
    planowanie 98  
    pojęcia 95  
    składniki 100  
    tworzenie 102  
przykładowy program ładujący,  
    uruchamianie 117

## S

serwer danych  
    definiowanie 19  
składniki EIP  
    administrowanie 4  
    Centrum informacyjne 7  
    eksploracja informacji 5  
    łączniki 5  
    przeglądarka danych 6  
    Web Crawler 5  
    wyszukiwanie obrazów 5  
    wyszukiwanie tekstowe 5

- składniki EIP (*kontynuacja*)
  - zgodność systemów operacyjnych 3, 4
- systematyka
  - korzystanie z narzędzia Information Structuring Tool 63
- szablony wyszukiwania, tworzenie 27

## U

- uruchamianie
  - program budujący przepływ pracy 102
- użytkownik 31
  - przenoszenie domen 40
  - zbiór uprawnień 36

## W

- Web Crawler
  - opcja EIP 5
- wyszukiwanie obrazów
  - konfiguracja 112
  - konfigurowanie 112
  - przypisywanie aliasu 114
  - weryfikacja połączenia 114
- wyszukiwanie tekstowe
  - konfiguracja 111
  - obsługa języka XML 111

## Z

- zasoby serwera 19
- zbiór uprawnień 31, 35, 36
  - przenoszenie domen 41
  - tworzenie 35
- zbiór uprawnień nadawania 36







Numer Programu: 5724-B43

Drukowane w USA

SC85-0067-01

