

Opis Techniczny

NetSprint Advanced Search Platform (NASP)

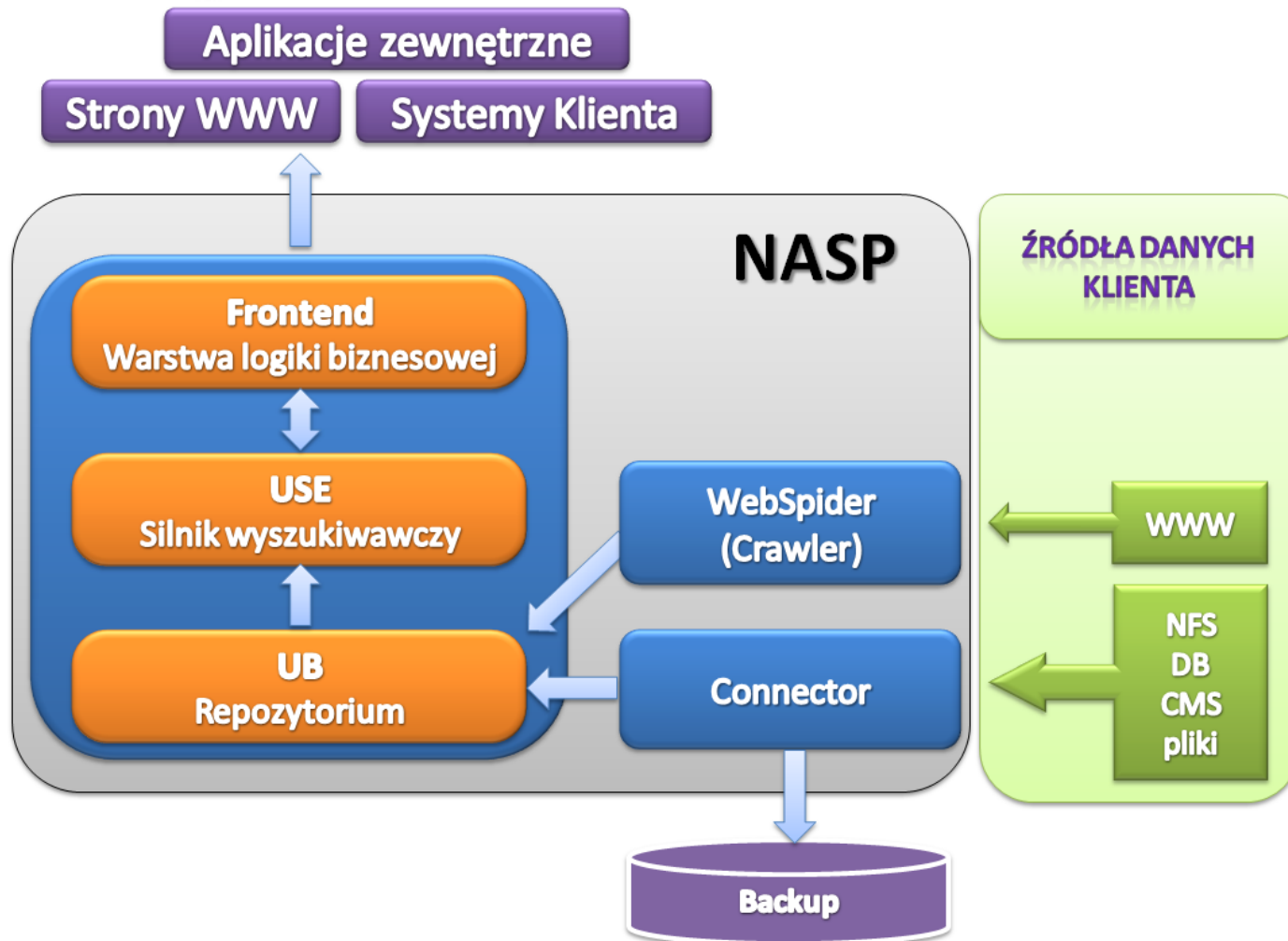
www.netsprint.eu

NASP – Uniwersalna platforma wyszukiwawcza

- Centralizacja danych pochodzących z różnych źródeł
- Zaawansowane scenariusze wyszukiwania
- Skalowalność i elastyczność
- Wysoka wydajność

Moduły dodatkowe umożliwiają:

- Otrzymanie wyników wyszukiwania podczas wpisywania zapytania
- Powiązanie dokumentów o podobnych treściach
- Dynamiczne generowanie stron tematycznych



Podstawowe elementy systemu:

- Connector lub Spider (web crawler)
- Repozytorium danych - UB
- Silnik wyszukiwawczy – USE
- Frontend – Business Logic Service

Moduły dodatkowe:

- AutoComplete
- Misspelling
- Related Content
- Topic Pages

Connector

- **Źródła danych** (system plików, baza danych, serwer FTP)
- **Format danych wejściowych** (XML, HTML, RSS, MSOffice, PDF)
- **Moduły Connectora**
 - Retriever
 - Converter
 - Validator
 - Buffer
 - Pusher

Repozytorium danych **UB (Universal Backend)**:

- Przechowuje dużą liczbę danych (TB)
- Umożliwia dodanie obsługi dowolnego typu pliku (wtyczki)
- Zapewnia szybki dostęp do danych
- Wspiera zarządzanie danymi (wstawianie, usuwanie)
- Zarządza procesem indeksowania danych
- Wspiera inne funkcjonalności operujące na danych, w tym:
 - Misspelling
 - AutoComplete
 - Related Content

Silnik wyszukiwawczy **USE (Search Engine)**:

Wyszukiwanie pełnotekstowe

- Słowa, frazy, konteksty
- Operatory logiczne (AND, OR, NOT)
- Sprowadzanie słów do ich form podstawowych (obsługa wielu form podstawowych) przy wykorzystaniu słowników i stemmerów
- Czułość na wielkość liter – możliwość konfiguracji

Wyszukiwanie po wartościach typu (w tym wyszukiwanie przedziałowe)

- numerycznego, tekstowego, daty (obsługa dat przed 01.01.1970)

Multiquery - łączenie wielu parametrów w zapytaniu

Silnik wyszukiwawczy USE (Search Engine):

Parametry filtrujące

- Zawężenia do wskazanej wartości konkretnego pola
- Możliwość łączenia wielu parametrów filtrujących jednocześnie

Sortowanie

- Ranking, pole numeryczne, mieszane, funkcja, łańcuchy, wiele kryteriów, ustalenie przedziału wartości sortowania, plugin-y

Zliczanie wyników w kategorii

Grupowanie (półzwijanie) – możliwość sortowania w obrębie grupy

Prezentacja wyników

- stronicowanie, przycinanie, podświetlanie

Frontend (Business Logic Service)

Zapewnia komunikację systemów Klienta z platformą wyszukiwawczą

- Zadawanie zapytań (protokół HTTP)
- Wyniki wyszukiwania (np. XML, HTML, JSON)

Umożliwia tworzenie rozbudowanych scenariuszy wyszukiwania

- Przetworzenie pojedynczego zapytania użytkownika na odpowiednie zapytania kierowanie do różnych wewnętrznych modułów systemu
- Połączenie odpowiedzi różnych mechanizmów w spójną odpowiedź przekazaną w wynikach wyszukiwania

Frontend (Business Logic Service)

Wyszukiwarka główna – Advanced Search i Federated Search

- Funkcjonalność Misspelling
- Odpowiednia liczba zapytań skierowana równolegle do indeksu silnika wyszukiwawczego USE

Wyszukiwarka szybka – Quick Search

- Funkcjonalność AutoComplete
- Szybki indeks silnika wyszukiwawczego USE

Auto Complete

Źródła danych (fraz):

- Słowniki
- Repozytorium (dokumenty znajdujące się w systemie)

Generowanie fraz

- Wstawienie do bazy danych fraz wraz z ich licznością

Indeksowanie fraz

- Wstawienie do indeksu fraz wraz z ich rankingiem

Zapytania do Auto Complete

Wyniki Auto Complete

Misspelling

Źródła danych (fraz):

- Słowniki
- Repozytorium (dokumenty znajdujące się w systemie)

Generowanie słów

- Wstawienie do bazy danych słów wraz z ich licznością

Indeksowanie słów

- Wstawienie do indeksu słów wraz z ich rankingiem

Zapytania do Misspelling

Wyniki Misspelling

Related Content – Moduł treści powiązanych

Źródła danych

- Repozytorium (dokumenty znajdujące się w systemie)

Ekstrakcja encji

- Generowanie encji na podstawie słownika

Keywordyzator

- Generowanie keywordów wraz ze statystykami wystąpień

Indeksowanie dokumentów wraz ze statystykami

Zapytania do Related Content

Wyniki Related Content

Topic Pages - Strony tematyczne

Dane statyczne

- Opis strony tematycznej

Dane dynamiczne

- Dokumenty ze zdefiniowanych źródeł
- Powiązane strony tematyczne
- Podobne strony tematyczne
- Bardziej szczegółowe / ogólne strony tematyczne

Topic Pages - Strony tematyczne

Należy zdefiniować

- Kategorie tematyczne
- Źródła danych
- Szablony

Panel Administracyjny - funkcjonalności

- Tworzenie nowych stron tematycznych
- Wprowadzanie opisu strony tematycznej
- Dowiązanie tematów powiązanych
- Definiowanie tematów bardziej ogólnych / szczegółowych

