

# POZYCJONOWANIE ZASOBÓW BIBLIOTEK CYFROWYCH

Adam Dudczak, **PCSS**

Tomasz Kalota, **Biblioteka Uniwersytetu Wrocławskiego**

II Konferencja Polskie Biblioteki Cyfrowe  
Poznań, 2009

# Wprowadzenie



- Po co pozycjonować?
- Kilka słów o założeniach pozycjonowania stron internetowych
  - ▣ ang. *Search engine optimization, SEO*
- Co możemy zrobić, a co inni zrobią za nas
- Kilka rzeczy o których należy pamiętać
- Ciemna strona mocy

# Po co pozycjonować?



- Scenariusz optymistyczny:
  - Jak odnaleźć igłę w stogu siana?
  - Zakładamy, że użytkownik wie czego szuka
- Scenariusz pesymistyczny:
  - Szukanie czegoś w stogu siana
  - Osoba szukająca nie wie nawet, że powinna szukać igły
  - Nie ma pojęcia o dziedzinie, nie zna nomenklatury, zwyczajów, konwencji

# Po co pozycjonować?



- Zadaniem pozycjonera jest sprowadzenie użytkowników na daną stronę
- W praktyce wystarczy wysokie miejsce na liście wyników wyszukiwania w Google/Yahoo/Bing
- W przypadku bibliotek cyfrowych
  - ▣ Widoczność w serwisach takich jak FBC, Europeana, OAlster ...

# Ogólne założenia SEO



- Odpowiedni dobór słów kluczowych
- Zgodność słów kluczowych w nagłówku meta z zawartością strony
- Poprawność techniczna strony
- Mapa strony
- Sensowny opis (atrybut „title”) i tekst w odnośniku

# Ogólne założenia SEO



- Rozróżnialne tytuły stron i ich zgodność z słowami kluczowymi
- Szybkie działanie strony
- Liczba i jakość linków do naszej strony z różnych witryn
- Dostępność danych w różnych formatach
- ...

# Podział obowiązków

- Dużo zależy od możliwości platformy użytej do budowy biblioteki cyfrowej
  - Poprawność techniczna strony
  - Właściwa konstrukcja odnośników
  - Rozróżnialne tytuły stron i ich zgodność z słowami kluczowymi
  - Zgodność nagłówka `meta` z zawartością strony
  - Szybkie działanie strony
  - Mapa strony
  - Dostępność danych w różnych formatach

# Podział obowiązków



- Często jest tak, że platforma spełnia pewne wymagania, ale potrzebna jest odpowiednia konfiguracja



# Podział obowiązków



- Co zatem zostaje dla twórców bibliotek cyfrowych?
  - Znajomość narzędzia którego używają
  - Odpowiedni dobór słów kluczowych
  - Dostępność tekstu publikacji (OCR)
  - Liczba linków prowadzących do naszej strony

# SEO katalogowanie

- Jak tworzyć „odnajdywalne” metadane?
- Wygląda na to, że wystarczy tworzyć dobre metadane
  - ▣ „Oh, co za ulga 😊”
- Świadomy wybór stosowanego schematu metadanych
  - ▣ Czy istnieje standard który idealnie pasuje do tego co jest nam potrzebne?
- Określenie i przestrzeganie konwencji opisu bibliograficznego

# Ocena jakości metadanych

- Metadane powinny informować użytkownika „o czym” jest obiekt
- Z punktu widzenia SEO:
  - ▣ Metadane powinny być obszerne
    - Uwzględniać różne wersje tego samego terminu
    - Uwzględniać formy pomyłkowe
  - ▣ Metadane powinny być zrozumiałe i przydatne
    - „Kto? Kiedy? Co? Gdzie?”

# Ocena jakości metadanych

- Z punktu widzenia SEO:
  - Zapewnienie interoperacyjności metadanych
    - Spójność na poziomie stosowanego słownictwa i konwencji zapisu np. dat, praw autorskich
    - Serwisy takie jak Europeana czy Driver oferują użytkownikom pewne dodatkowe funkcje, ale stawiają też wymagania
      - np. Europeana timeline

# Formy omyłkowe



- Gdzie jest miejsce dla form omyłkowych?
- Czy używać dLibrowego słownika synonimów?
  - Łatanie opisu bibliograficznego
- Należy pamiętać, że tylko to co jest w metadanych zostanie zaindeksowane przez Google czy FBC

# Metadane raz na zawsze?

- W komercyjnym SEO utrzymanie wysokiego miejsca w wynikach wyszukiwania to nie zadanie jednorazowe
- Konkurencja, konkurencja, konkurencja!
- Zmieniają się trendy, zmienia się język, zmienia się informacja
- Będą musiały zmienić się również metadane
- Czy takie działania są podejmowane w tej chwili?
  - Czasem tak, np. aktualizacja informacji biograficznego po śmierci autora

# Metadane raz na zawsze?

- Tagi wprowadzane przez użytkowników to swoisty wentyl bezpieczeństwa
- Zmiany, które są konieczne z punktu widzenia nowych funkcji, np.
  - ▣ wprowadzenie interfejsu prezentującego zasoby na mapie

# Automatyczna analiza metadanych



- Konieczność dokonania masowych zmian w istniejących metadanych
  - ▣ np. przyłączanie do Europeany, Driver'a
- Co można sprawdzić automatycznie?
  - ▣ Ocena kompletności opisu względem zastosowanego schematu
  - ▣ Spójność stosowanego słownictwa
    - Na podstawie predefiniowanego słownika
    - Analiza statystyczna



# Automatyczna analiza metadanych



- Co można sprawdzić automatycznie?
  - Zgodność stosowanego słownictwa z tym wykorzystywanym przez użytkowników do wyszukiwania
  - Zgodność użytych słów z treścią publikacji (gdy dostępny jest OCR)
- Wyniki automatycznej analizy jako wskazówka nie wyrocznia

# Granularność obiektów



- Czy publikować całe czasopismo jako jedną publikację, czy każdy artykuł z osobna?
- Podział daje szansę na stworzenie dokładniejszych metadanych dla pojedynczego artykułu
- Dobrze z punktu widzenia dostępności – mniejsze pliki
  - ▣ Ważne w przypadku urządzeń mobilnych

# Wielojęzyczność

- W tej chwili z zasobów polskich bibliotek cyfrowych korzystają głównie Polacy
- Czy osoba nie znająca języka polskiego jest w stanie odnaleźć nasze zasoby?
- Opisy w różnych językach to zdecydowanie dobra praktyka
- Stosowanie standardowych słowników kontrolowanych umożliwiłoby automatyczne tłumaczenie metadanych
- Tłumaczenie maszynowe – Google Translate

# Dobry OCR na wagę złota!

- Metadane to „tylko” metadane
- Ważne jest żeby udostępnić treść obiektów dla robotów wyszukiwarek
- Stosowanie odpowiedniego oprogramowania OCR
  - wsparcie dla języka rozpoznawanego tekstu
- Dobór odpowiedniego formatu obiektów cyfrowych
  - Roboty wyszukiwarek muszą ten format rozumieć

# Linki do strony biblioteki cyfrowej

- W jaki sposób pozyskiwać?
- Dodawanie linków do zasobów BC w innych usługach:
  - Wikipedia – łączenie z powiązаныmi tematycznie hasłami – wydarzeniami/ludźmi/miejscami
- Blogi typu „bibliotekarz cyfrowy poleca”
- Biblioteki (zbiory?) odnośników
  - HiNT (<http://www.epublica.com.pl/hint>)

# Linki do strony biblioteki cyfrowej

- Wzajemne linki między katalogiem, a biblioteką cyfrową
  - ▣ Wyciąganie katalogów z czeluści ukrytego internetu
  - ▣ Mały zysk dla BC, duży dla katalogów i użytkowników
- Komponent wyświetlający podobne publikacje z FBC w interfejsie BC
- Baza projektów digitalizacyjnych FBC
  - ▣ Linki do polecanych zasobów

# Udostępnianie metadanych

- Łatwiej wejść do biblioteki, która ma kilkoro drzwi wejściowych
- Dziedzione i ogólne usługi agregujące
  - OAlster, ScientificCommons, FBC, Europeana
  - Manuscriptorium (<http://manuscriptorium.com>)
- Interoperacyjność metadanych warunkiem pełnej integracji z innymi usługami
  - lepsza widoczność zasobów
  - np. Europeana Timeline, Europeana Faceted Search

# Udostępnianie treści obiektów



- Również agregacja treści obiektów cyfrowych może się przysłużyć promocji zasobów
  - np. Flickr Commons
  - Obiekty w niższej jakości, link do pełnej wersji zasobu



# Tunnel, Wombeyan Road

ALL SIZES



**Format:** Glass plate negative.

**Rights Info:** No known restrictions on publication.

**Repository:** Tyrrell Photographic Collection, Powerhouse Museum  
[www.powerhousemuseum.com/collection/database/collection=The\\_Tyrrell\\_Photographic](http://www.powerhousemuseum.com/collection/database/collection=The_Tyrrell_Photographic)

**Part Of:** Powerhouse Museum Collection

ph<sup>™</sup> Uploaded on October 12, 2008  
by [Powerhouse Museum Collection](#)

## Powerhouse Museum Collection's photostream

1,398 uploads

browse

This photo also belongs to:

## Tyrrell Collection (Set)

1,005 items

browse

5 people call this photo a favorite

## Tags

- Powerhouse Museum
- B&W
- Tunnel
- Trees
- Car
- wombeyan
- mittagong

# Ciemne strony SEO

- Czy użytkownik nie będzie czuł się okłamany?
- Nie chodzi nam o złośliwe sztuczki, ale o podkreślenie tego co jest istotne w kontekście funkcjonowania BC w sieci
- Być może warto spisać te dobre praktyki i umieścić
  - ▣ np. w „E-poradniku redaktora zasobów cyfrowych”
- Dlaczego warto dbać o widoczność polski

# Podsumowanie

---

- Jeszcze więcej pracy dla bibliotekarzy cyfrowych  
😊
- Być może warto spisać te dobre praktyki
  - ▣ np. umieścić w „E-poradniku redaktora zasobów cyfrowych” w formie listy spraw do sprawdzenia
- Dlaczego warto dbać o widoczność polskich zasobów w sieci?

# DZIĘKUJEMY ZA UWAGĘ

Adam Dudczak, [maneo@man.poznan.pl](mailto:maneo@man.poznan.pl)

Tomasz Kalota, [tomek@kalota.pl](mailto:tomek@kalota.pl)