

## MOŻLIWOŚCI WYSZUKIWANIA DOKUMENTÓW MUZYCZNYCH UTWORZONYCH ZGODNIE Z ARCHITEKTURĄ IODA

Adam Łukasz KACZMAREK

Politechnika Gdańska; Wydział Elektroniki, Telekomunikacji i Informatyki  
tel: 58 347 13 78 e-mail: adam.l.kaczmarek@eti.pg.gda.pl

**Streszczenie:** Artykuł przedstawia zagadnienie zastosowania dokumentów muzycznych utworzonych zgodnie z architekturą IODA (ang. Interactive Open Document Architecture). Dokumenty w architekturze IODA składają się z wielu plików powiązanych ze sobą semantycznie. Zależności te definiowane są w tzw. grzbiecie (ang. spine) dokumentu będącym plikiem w formacie XML (ang. eXtensible Markup Language). Dokumenty muzyczne tworzone zgodnie z architekturą IODA zawierają ścieżki dźwiękowe utworów, słowa piosenek, zapis nutowy melodii i inne rodzaje danych. Zastosowanie dokumentów muzycznych IODA rozszerza możliwości wyszukiwania utworów muzycznych. Dzięki architekturze IODA utwory muzyczne można wyszukiwać między innymi na podstawie fragmentów tekstów śpiewanych w utworach oraz na podstawie zapytań QbH (Query by Humming), które są fragmentami melodii występującymi w utworach.

**Słowa kluczowe:** wyszukiwanie informacji, zapytania, dokumenty muzyczne, architektura dokumentów.

### 1. WPROWADZENIE

Sieć Internet pozwala na pozyskanie informacji i danych zapisanych w wielu różnych formatach. Dane oznaczają zarejestrowane fakty, natomiast informacje są produktem przetwarzania danych pozwalających na ich interpretację. Informacje i dane są rozpowszechniane między innymi w postaci tekstu, obrazków, strumieni wideo oraz ścieżek dźwiękowych. Zasadniczym problemem podczas korzystania z Internetu jest konieczność odszukania informacji, którymi użytkownik jest zainteresowany. W najszerszym zakresie metody wyszukiwania zostały zaadaptowane do informacji i danych zapisanych w postaci tekstu. Powszechnie stosowane wyszukiwarki internetowe przeprowadzają wyszukiwanie przede wszystkim słów występujących na stronach internetowych i w innych dokumentach udostępnionych w sieci. Jednak istnieje również możliwość wyszukiwania dokumentów graficznych na podstawie ich zawartości oraz plików dźwiękowych na podstawie sekwencji dźwięków, jakie w nich występują. Niniejszy artykuł poświęcony jest plikom dźwiękowym.

Artykuł przedstawia zastosowanie architektury IODA (ang. Interactive Open Document Architecture) do usprawnienia procesu wyszukiwania danych dźwiękowych. Za pomocą tej architektury pliki i dane dotyczące dźwięku zapisywane są w postaci tzw. dokumentów muzycznych. Dokumentem muzycznym jest to dokument dźwiękowy, którego główną treść stanowi muzyka. Treścią dokumentów dźwiękowych mogą być natomiast inne sekwencje dźwięków. Dokumenty muzyczne w architekturze IODA

składają się z wielu plików powiązanych ze sobą semantycznie. Powiązania semantyczne zapisywane są w postaci pliku XML (ang. Extensible Markup Language). Zorganizowanie danych w postaci tego rodzaju dokumentów pozwala na rozszerzenie możliwości wyszukiwania informacji i danych dźwiękowych.

### 2. WYSZUKIWANIE PLIKÓW DŹWIĘKOWYCH

Podczas wyszukiwania dowolnego rodzaju informacji i danych stosuje się zapytania (ang. query). Zapytanie jest zbiorem danych, które określa rodzaj poszukiwanych zasobów. W przypadku wyszukiwania plików muzycznych zapytanie może mieć kilka różnych postaci. Jednym z najprostszych rodzajów zapytań do danych muzycznych to takie, które zawiera tytuł, autora utworu muzycznego lub inne metadane. Metadane są to dane opisujące treść pewnego zbioru danych. Metadanymi jest na przykład rok skomponowania utworu oraz nazwa albumu zawierającego utwór. Posiadanie takich danych w ogólności pozwala na szybkie odszukanie utworu muzycznego lub informacji o nim. Użytkownik poszukujący utworu może jednak nie znać jego tytułu, autora ani innych danych o utworze. Wyszukiwanie jest również możliwe, jeśli użytkownik zna fragment tekstu występujący w utworze lub fragment melodii tego utworu.

Utwór muzyczny można znaleźć na podstawie fragmentu tekstu, jaki jest w nim zawarty w taki sposób, że najpierw na podstawie znanego fragmentu określa się tytuł utworu oraz jego autora, a następnie na podstawie tych metadanych wyszukuje się ścieżkę dźwiękową utworu. Określić tytuł i autora utworu na podstawie fragmentu tekstu zawartego w utworze można za pomocą wyszukiwarek internetowych ogólnego przeznaczenia. Podanie znanej frazy, jako zapytania do wyszukiwarki wraz z dodatkowymi informacjami wskazującymi na to, że poszukiwany jest tekst piosenki, pozwala zawęzić obszar poszukiwań do zbioru stron internetowych zawierających informacje o wyszukiwanym utworze.

Inny rodzaj wyszukiwania występuje wtedy, gdy użytkownik poszukuje piosenki na podstawie fragmentu melodii zawartej w tej piosence. Zapytanie zawierające taki fragment melodii nazywane jest zapytaniem QbH (Query by Humming) [1]. Istnieje szereg serwisów internetowych przeznaczonych do realizacji takich zapytań. Opracowane zostały również liczne algorytmy przeznaczone do realizacji tego rodzaju zapytań [1][2]. Niestety, wyniki wyszukiwań przeprowadzanych za pomocą takich zapytań zawierają

liczne błędy. Dopasowanie zapytania do melodii zawartych w plikach muzycznych jest problematyczne. Wynika to z faktu, że na utwory muzyczne składa się wiele różnych dźwięków wytwarzanych za pomocą różnych instrumentów muzycznych oraz ludzkiego głosu. Wyodrębnienie samej melodii z utworu muzycznego wymaga przetworzenia tego pliku w szerokim zakresie. Problem wyszukiwania za pomocą zapytań typu QbH można jednak rozwiązać za pomocą wykorzystaniu architektury IODA.

Mechanizmy wyszukiwania na podstawie zapytań typu QbH oraz na podstawie innych zapytań odnoszących się do zawartości utworu muzycznego są częściowo wspierane przez standardy MPEG7 i MPEG21 przeznaczone do zapisu plików muzycznych [3]. Pliki zapisane zgodnie z tymi standardami mogą zawierać tzw. deskryptory (ang. descriptors), które określają dane dotyczące zawartości utworów i przez to wspomagają ich wyszukiwanie.

### 3. ARCHITEKTURA IODA

Architektura IODA przeznaczona jest do tworzenia dokumentów elektronicznych posiadających cztery następujące cechy [4]:

- wykonywalność,
- mobilność,
- inteligencję,
- interaktywność.

Wykonywalność dokumentu polega na tym, że oprócz tekstu czy informacji w postaci graficznej, dokumenty w architekturze IODA posiadają również funkcjonalność. Na przykład mogą one uruchamiać się na urządzeniu użytkownika w różny sposób, w zależności od rodzaju urządzenia. Ponadto w ramach dokumentu mogą być uruchamiane programy komputerowe. Na przykład artykuł naukowy może zawierać skrypt pozwalający na wykonanie obliczeń opisanych w tym artykule.

Inną cechą dokumentu w architekturze IODA jest mobilność [5]. Dokument posiada zdolność do przemieszczania się po sieci Internet. Dokument zawiera może listę adresatów, do których jest on dostarczany. Adresaci ci mogą mieć za zadanie zarówno odczytać dane z dokumentu, jak i pewne dane do niego wprowadzić.

Dokumenty w architekturze IODA posiadają również taką cechę, że są inteligentne. Cecha ta polega to na tym, że dokumenty funkcjonują jak agenty komputerowe w systemie wieloagentowym i mogą między innymi wymieniać między sobą dane [6].

Ostatnią cechą charakterystyczną dokumentów w architekturze IODA jest interaktywność. Polega ona na współpracy z użytkownikiem w dostarczaniu i wydobywaniu danych. Na przykład dokument może prezentować użytkownikowi sugestie dotyczące tego, jakie dane użytkownik powinien wprowadzić, zachowując przy tym prawo użytkownika do udostępnienia tych danych [7].

Dokumenty utworzone w architekturze IODA charakteryzują się tym, że składać się mogą z wielu plików. Wśród tych plików mogą być pliki tekstowe, wykonywalne pliki binarne, skrypty itp. Wszystkie te pliki połączone są ze sobą za pomocą pliku w formacie XML, który stanowi grzbiet (ang. spine) dokumentu. Nazwa pochodzi od łączenia wielu zapisanych kartek papieru w jeden dokument za pomocą plastikowej oprawy. Plik ten zawiera dane o semantycznych relacjach między różnymi plikami wchodzącymi w skład dokumentu IODA. Plik taki określa również powiązania pomiędzy różnymi fragmentami

dokumentów. Na przykład definiować może uruchomienie pewnego skryptu w określonej części dokumentu. Sposób tworzenia dokumentów za pomocą grzbietu w formacie XML porównać można do strony internetowej, w której główną część stanowi plik tekstowy w postaci pliku HTML, w którym są odnośniki do innych plików, takich jak pliki graficzne, które wchodzą w skład strony internetowej. W tym przypadku również jeden dokument, którym jest strona internetowa składa się z wielu plików.

### 4. DOKUMENTY MUZYCZNE IODA

Zastosowanie architektury IODA do danych muzycznych prowadzi do utworzenia dokumentów muzycznych. Zostały one nazwane IMD (ang. IODA Music Documents) [8]. Dokument taki zawiera powiązane ze sobą semantycznie pliki zawierające dane dotyczące utworów muzycznych. Dokument muzyczny IMD zachowuje wszystkie cechy dokumentu IODA, jakimi są wykonywalność, mobilność, inteligencja i interaktywność.

W skład dokumentu muzycznego IMD wchodzić mogą następujące rodzaje plików:

- ścieżka dźwiękowa utworu (np. pliki w formatach MP3, OGG, FLAC),
- słowa utworu muzycznego (np. pliki w formatach TXT, HTML, PDF),
- wersja utworu w postaci pliku MIDI,
- metadane,
- strumień video (np. pliki w formatach AVI, MP3),
- inne pliki.

Powiązania pomiędzy plikami określane są w grzbiecie dokumentu będącym plikiem w formacie XML. Powiązania między plikami odnoszą się mogą do całych plików, jak również do ich fragmentów. Na przykład powiązanie między dwoma plikami zapisane w grzbiecie dokumentu polegać może na tym, że dany jest plik ze ścieżką dźwiękową utworu oraz plik z tekstem tego utworu. Zależności semantyczne zapisywane w grzbiecie mogą być również bardziej szczegółowe. Mianowicie, plik XML zawierać może dane dotyczące tego, które fragmenty tekstu piosenki są śpiewane w poszczególnych chwilach czasu trwania utworu. Podobna relacja, jak między ścieżką dźwiękową utworu a tekstem piosenki, zachodzić może między ścieżką dźwiękową a melodią piosenki zapisaną w pliku o formacie MIDI. Pliki takie zawierają tylko i wyłącznie sekwencję ściśle określonych dźwięków. Pliki MIDI zawierać mogą sekwencję dźwięków będących melodią utworu muzycznego. Informacje o takiej zależności również mogą być zapisywane w grzbiecie dokumentu muzycznego.

Informacje i dane zapisywane w pliku XML nie powodują modyfikowania plików wchodzących w skład dokumentu. Jest to bardzo istotna cecha dokumentów IMD. Pozwala to na tworzenie dokumentów muzycznych z istniejących plików. Dodatkowo istniejące aplikacje komputerowe przeznaczone do przetwarzania plików dźwiękowych mogą korzystać z plików wchodzących w skład dokumentów muzycznych niezależnie od tego, że pliki te są elementem dokumentów.

Plik XML będący grzbieciem dokumentu IMD definiuje sekwencję zależności pomiędzy dwoma punktami zakotwiczenia (ang. anchoring). Każdy z tych punktów jest albo całym plikiem albo fragmentem pliku albo pewną lokalizacją w pliku wchodzącym w skład dokumentu muzycznego. Istnieje konieczność różnego sposobu definiowania punktów zakotwiczenia w zależności od tego,

w jakim pliku występują. Na przykład w przypadku plików tekstowych punktem zakotwiczenia może być pewien znak znajdujący się w określonym miejscu pewnego wiersza pliku. W pliku XML dokumentu IMD dane takie umieszcza się zgodnie ze specyfikacją plików XML za pomocą znaczników.

W przypadku dokumentów IMD zastosować można znaczniki o następujących nazwach <imd:line> i <imd:character>. Dla plików ze ścieżką dźwiękową utworów punktem zakotwiczenia może być pewien czas, jaki upłynął od momentu rozpoczęcia utworu. W dokumentach IMD mogą zostać umieszczone znaczniki <imd:hour>, <imd:minute> and <imd:second> definiujące kolejno liczbę godzin, minut i sekund, jakie upłynęły od rozpoczęcia utworu. Dla plików będących obrazkami punktem zakotwiczenia mogą być współrzędne pewnego punktu obrazka. Punkty te określić można za pomocą znaczników <imd:dimensionx> and <imd:dimensiony>. Przykładowo fragment pliku XML definiującego dokument muzyczny może mieć następującą zawartość:

```
<ioda:link>
  <ioda:point>
    <imd:file> song.mp3 </imd:file>
    <imd:minute> 0 </imd:minute>
    <imd:second> 30 </imd:second>
  <ioda:point>
  <ioda:point>
    <imd:file> lyrics.txt <imd:file>
    <imd:line> 3 <imd:line>
    <imd:character> 5 <imd:character>
  <ioda:point>
</ioda:link>
```

Fragment ten określa powiązanie pomiędzy pewną chwilą czasu w pliku song.mp3, a pewnym miejscem w pliku tekstowym lyrics.txt.

Taki sposób definiowania punktów zakotwieczeń posiada jednak istotną wadę polegającą na tym, że modyfikacja pliku może spowodować, że punkt zakotwiczenia się zmieni. Problem ten dotyczy przede wszystkim zakotwieczeń w plikach tekstowych. Jeśli początek pliku tekstowego zostanie zmodyfikowany, a plik XML określający zakotwiczenie nie zostanie zaktualizowany, to zakotwiczenie stanie się nieprawidłowe.

Problem ten można rozwiązać, stosując metody określania zakotwieczeń znane z technologii dodawania przypisów (ang. annotations) do dokumentów [9]. Przypisy te są to dodatkowe notatki, jakie użytkownik może umieścić w czytany przez niego dokumencie. Technologia ta stosowana jest na przykład w dokumentach w formacie PDF. Przypisy dodawane są w taki sposób, że zostają powiązane z pewnym fragmentem pliku, na przykład sekwencją wyrazów lub zdań. Jeśli sekwencja ta nie zostanie zmodyfikowana, zakotwiczenie pozostaje prawidłowe nawet wtedy, gdy zostaną wprowadzone zmiany w innych częściach dokumentu.

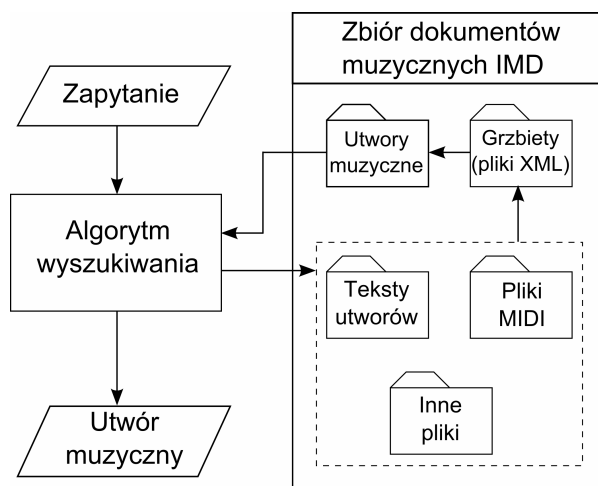
Plik XML oprócz punktów zakotwiczenia określa również lokalizację plików, w których punkty te występują. Standard IODA umożliwia umieszczanie plików wchodzących w skład dokumentu w różnych lokalizacjach. Wymienione pliki mogą znajdować się na tym samym komputerze, na którym znajduje się grzbiet dokumentu będący plikiem XML. Mogą być umieszczone w tym samym katalogu, co grzbiet dokumentu, jak również w innych katalogach. Jeśli pliki umieszczane są w innych katalogach, plik XML musi zawierać ścieżkę do powiązanych ze sobą

plików. Dodatkowo standard IODA pozwala na tworzenie zależności między plikami, które dostępne są zdalnie za pośrednictwem sieci Internet. W celu tworzenia zależności między takimi zasobami stosuje się powszechnie używane metody adresowania zasobów, tj. format URL (ang. Uniform Resource Locator).

## 5. ZASTOSOWANIE DOKUMENTÓW IMD

Jednym z najważniejszych zastosowań dokumentów muzycznych IMD jest rozszerzenie możliwości wyszukiwania plików muzycznych. Tworzenie kolekcji utworów muzycznych w postaci dokumentów IMD pozwala na przeprowadzanie wyszukiwania utworów za pomocą metadanych tj. tytuł utworu, czy jego autor. Architektura IODA dodatkowo zwiększa możliwości przeprowadzania wyszukiwania na podstawie fragmentu tekstu utworu oraz zapytań QbH. Wyszukiwanie takie przeprowadzane może być bez konieczności wykorzystania zasobów zgromadzonych w sieci Internet. Użytkownik może mieć lokalnie zapisany zbiór utworów muzycznych, wśród których przeprowadzać może wyszukiwanie, korzystając z wyżej wymienionych metod.

Schemat wyszukiwania utworów muzycznych zgromadzonych w kolekcji utworów zapisanych w postaci dokumentów IMD przedstawiony jest na rysunku 1.



Rys. 1. Wyszukiwanie utworów muzycznych zgromadzonych w postaci dokumentów IMD

Rysunek ten uwzględnia trzy rodzaje zapytań. Są to zapytania zawierające:

- metadane,
- fragment utworu,
- melodię QbH.

W przypadku wyszukiwania za pomocą fragmentu utworu oraz zapytań QbH nie istnieje konieczność dopasowywania zapytań do ścieżek dźwiękowych utworów muzycznych. W przypadku zapytań będących fragmentem tekstu utworu dopasowywania takie byłoby możliwe w sytuacji przeprowadzenia procesu rozpoznawania mowy śpiewanej w utworze muzycznym, a następnie dopasowanie zapytania do rozpoznanego tekstu. Jakość takiego wyszukiwania byłaby niska, ponieważ występowałyby liczne błędy wynikające z trudności rozpoznawania tekstu śpiewanego w piosenkach. Korzystając natomiast z architektury IODA, zapytanie będące fragmentem utworu dopasowywane jest do pliku tekstowego zawierającego tekst utworu. Plik z tekstem utworu powiązany jest za pomocą

grzbietu dokumentu z plikiem zawierającym ścieżkę dźwiękową utworu. Jeśli nastąpi dopasowanie pomiędzy zapytaniem użytkownika a tekstem zawartym w pliku tekstowym, wtedy system wyszukiwujący może na podstawie pliku tekstowego określić, jaki utwór muzyczny zawiera dany tekst.

Podobny schemat wyszukiwania występuje w sytuacji wyszukiwania za pomocą zapytań QbH. W przypadku wykorzystania dokumentów IMD wyszukiwanie zapytaniami typu QbH nie musi odbywać się na plikach zawierających utwór muzyczny, lecz wyszukiwanie przeprowadzane jest wśród plików MIDI, które są odpowiednikami utworów muzycznych. Pliki MIDI są sekwencjami ściśle zdefiniowanych dźwięków, w związku z tym dopasowanie zapytania QbH do takich plików jest mniej problematyczne niż dopasowywanie zapytań QbH do plików z utworami muzycznymi [10]. Po dopasowaniu zapytania QbH do pliku MIDI możliwe jest wydobycie utworu muzycznego, któremu dany plik MIDI odpowiada, dzięki grzbietowi dokumentu, w jakim zapisane jest to powiązanie.

## 6. ZAKOŃCZENIE

W celu zrealizowania wyżej opisanego wyszukiwania konieczne jest posiadanie aplikacji korzystającej z możliwości IMD. Aplikacja taka może być uruchamiana lokalnie na urządzeniu użytkownika, na którym zgromadzony jest zbiór utworów muzycznych w postaci dokumentów IMD. Dodatkowo przeprowadzenie wyszukiwania wymaga posiadania zbioru tego rodzaju dokumentów, zawierających oprócz utworów muzycznych również pliki z nimi powiązane, w szczególności pliki MIDI i pliki z tekstem utworów. Kolekcja taka utworzona może zostać ręcznie, jak również możliwe jest opracowanie aplikacji wspierających tworzenie dokumentów IMD na podstawie zbioru plików z danymi muzycznymi. W przypadku rozpowszechnienia standardu IODA utwory muzyczne mogłyby być dostarczane razem z dodatkowymi plikami powiązanymi z nimi semantycznie wspierającymi wyszukiwanie utworów.

## 7. BIBLIOGRAFIA

1. Kotsifakos A., Papapetrou P., Hollmén J., Gunopulos D., Athitsos V.: A Survey of Query-by-humming

- Similarity Methods, Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '12, Heraklion, Crete, Greece, ACM, 2012, s. 5:1-5:4.
2. Zhu Y., Shasha D.: Warping Indexes with Envelope Transforms for Query by Humming, Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, SIGMOD '03, San Diego, California, ACM, 2003, s. 181-192.
3. Batke J.-M., Eisenberg G., Sikora T., Weishaupt P.: A Query by Humming System using MPEG-7 Descriptors, Proceedings of the 116<sup>th</sup> AES Convention, Audio Engineering Society, New York, 2004.
4. Siciarek J., Wiszniewski B.: IODA - an Interactive Open Document Architecture, Proceedings of the Int. Conf. on Computational Science, ICCS 2011, Procedia Computer Science, Nr 4 (0), Elsevier, 2011, s. 668-677.
5. Godlewska M., Wiszniewski B.: Distributed MIND - A New Processing Model Based on Mobile Interactive Documents., Parallel Processing and Applied Mathematics., Lecture Notes in Computer Science Nr 6068, Springer Berlin / Heidelberg, 2010, s. 244-249.
6. Godlewska M.: Agent System for Managing Distributed Mobile Interactive Documents, Agent and MultiAgent Systems: Technologies and Applications, LNCS, Nr. 6071, Springer Berlin / Heidelberg, 2010, s. 390-399.
7. Kaczmarek A.L.: Interactive Query Expansion With the Use of Clustering-by-Directions Algorithm, IEEE Transactions on Industrial Electronics, Nr 58, 2011, s. 3168-3173
8. Kaczmarek A.L.: The Application of the IODA Document Architecture to Music Data, Proc. of the 6th Int. Conf. on Knowledge Management and Information Sharing, INSTICC, Rome, Italy, 2014, s. 268-273.
9. Sevcech J., Bielikova M.: Query Construction for Related Document Search Based on User Annotations, Federated Conference on Computer Science and Information Systems (FedC-SIS), PTI, Polish Information Processing Society, Kraków, Poland, 2013, s. 279-286.
10. Francu C., Nevill-Manning C.: Distance Metrics and Indexing Strategies for a Digital Library of Popular Music, IEEE International Conference on Multimedia and Expo, ICME, Nr 2, IEEE, New York, USA, 2000, s. 889-892.

## THE USAGE OF MUSIC DOCUMENTS CREATED IN THE IODA ARCHITECTURE

The paper presents the usage of music documents created with respect to the IODA (Interactive Open Document Architecture) architecture. Documents in the IODA architecture consist of many files which are semantically related to each other. Relations are defined in the spine of the document. The spine is a file in the XML (eXtensible Markup Language) format. Music documents created in the IODA architecture contain music tracks, lyrics of songs, musical notations and other kind of data. Taking advantage of IODA music documents extends possibilities of performing the search for music data. The IODA architecture supports the search for songs on the basis of fragments of lyrics and on the basis of queries by humming (QbH) which are fragments of melodies occurring in songs.

**Keywords:** information retrieval, queries, music documents, document architecture.