

OpenDocument na biurku

# CO SIEDZI W ŚRODKU?

Twórcy formatu OpenDocument proponują nowe podejście do tematu wymiany i przechowywania danych przez aplikacje biurowe. Co to oznacza dla użytkownika? Co kryje się w środku pliku ODF?

Na ile przenośny jest OpenDocument?

W bieżącym temacie miesiąca postaramy się znaleźć odpowiedzi na wszystkie te pytania.

**DMITRI POPOV**

**W** ubiegłym roku stan Massachusetts przyjął otwarty format OpenDocument jako standard dla aplikacji biurowych. Na temat tego, co się tam wtedy działo, można by napisać niezłą powieść. Trudno uwierzyć, że zwykły format pliku wywołał tyle kontrowersji i stał się przedmiotem tylu debat politycznych. OpenDocument nie jest jednak zwykłym formatem dokumentu. Za skrótem ODF (Open Document Format) kryje się nowe spojrzenie na problem wymiany danych między aplikacjami biurowymi. Format OpenDocument zwiastuje koniec długiego rozdziału historii, w którym pro-

ducenci oprogramowania mogli kontrolować rynek narzędzi biurowych, zamykając dostęp do informacji na temat stosowanego formatu zapisu danych.

W tej batalii nie chodzi jednak ani o podgryzanie Microsoftu, ani o promowanie pakietu OpenOffice.org. Nie chodzi też o wszczynanie sporu na temat technicznej wyższości formatu Microsoft Word lub OpenOffice Writer. Przejście na otwarte standardy – takie jak ODF – pozwoli uniknąć zakłóceń przepływu informacji, powodowanych często przez korzystanie z zastrzeżonych, korporacyjnych formatów.



## Listing 1: content.xml

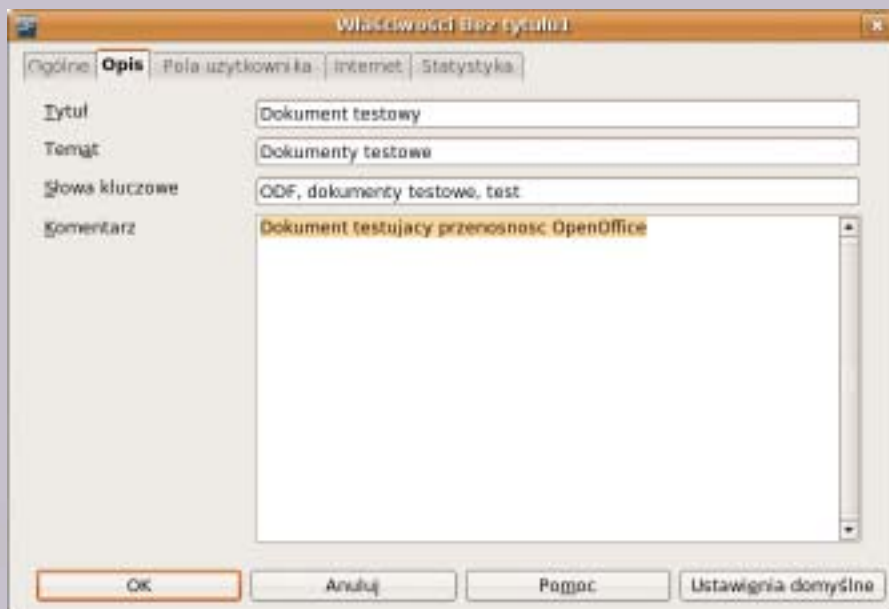
```
<text: h text: style-name=„Heading_20_1”text:outline-level=„1”>Lorem ipsum</text:h>
<text: p text: style-name=„Standard”/><text:p text:style-name=„Standard”>Lorem ipsum dolor sit amet. <text:span
text:style-name=„T1”>Proin velit.</text:span>
```

## Listing 2: manifest.xml

```
<? xml version="1.0" encoding="UTF-8"?>
<manifest:manifest xmlns:manifest="urn:oasis:names:tc:opendocument:xmlns:manifest:1.0">
  <manifest:file-entry manifest:media-type="application/vnd.oasis.opendocument.text" manifest:full-path="/" />
  <manifest:file-entry manifest:media-type="text/xml" manifest:full-path="content.xml" />
  <manifest:file-entry manifest:media-type="text/xml" manifest:full-path="styles.xml" />
  <manifest:file-entry manifest:media-type="text/xml" manifest:full-path="meta.xml" />
  <manifest:file-entry manifest:media-type="text/xml" manifest:full-path="settings.xml" />
</manifest:manifest>
```

OpenDocument Format to standard opracowany przez organizację OASIS (Organization for the Advancement of Structured Information Standards). Pierwotnym celem powstania formatu była standaryzacja wykorzystania XML w roli formatu zapisu i wymiany danych między aplikacjami biurowymi. Narodziny ODF już teraz zmusiły producentów zamkniętych rozwiązań, między innymi Microsoft, do zapewnienia lepszej alternatywy do dotychczasowych formatów zapisu danych. Ostatnio hasło OpenDocument często pojawia się w mediach a od premiery OpenOffice 2.0 użytkownicy na całym świecie stosują go w codziennej pracy. O samym formacie nie napisano jednak dotychczas zbyt wiele. W tym miesiącu przedstawimy, jak ODF wygląda od kuchni, i jak wygląda kwestia jego zgodności z alternatywnymi pakietami biurowymi. W tym artykule pokażemy jak dobrać się do wnętrza dokumentu ODF i omówimy sposób organizacji treści. W kolejnych, zajmiemy się kwestią przenośności formatu i opiszemy, co zrobić z plikiem ODF, aby bez niespodzianek otworzył się w programie Microsoft Word.

Korzystając z okazji, rzucimy też okiem na pakiety biurowe alternatywne względem OpenOffice. Poszukamy też przyczyn ostatniej mody na aplikacje oparte na WWW. Mamy nadzieję, że temat formatu OpenDocument i linuxowych aplikacji biurowych zainteresuje naszych Czytelników. Bardziej



Rysunek 1: Dużą część informacji z pliku meta.xml można podejrzeć w oknie właściwości dokumentu OpenOffice 2.0.

wyczerpujące informacje o strukturze OpenDocument można znaleźć w darmowej, opublikowanej na licencji GNU książce w formacie PDF pod tytułem *OpenDocument Essentials* autorstwa J. Davida Eisenberga, dostępnej pod adresem [2].

### Bogate wnętrze

Biurowe dokumenty nie składają się jedynie z tekstu i danych liczbowych. Typowy dokument zawiera również informacje o formato-

waniu, autorze dokumentu – i jeszcze wiele, wiele innych. ODF, uważany zazwyczaj za jeden plik, to w rzeczywistości cały ich zestaw. Na plik OpenDocument składa się kolekcja plików zapisanych w postaci Java Archive (JAR), czyli odmiany archiwum zip.

Aby zajrzeć do wnętrza dokumentu ODT, wystarczy zmienić jego rozszerzenie na `.zip` i rozpakować. Archiwum zawiera zwykle co najmniej następujące pliki:

- content.xml
- META-INF/manifest.xml
- meta.xml
- mimetype
- settings.xml
- styles.xml

Szczegóły mogą różnić się w zależności od rodzaju dokumentu – jeżeli na przykład dokument zawiera obrazy i makra, elementy te zostaną oddzielnie zapisane w dodatkowych katalogach.

## Listing 3: meta.xml

```
<meta:generator>OpenOffice.org/2.0$Unix OpenOffice.org_project/680m5$Build-9073</meta:generator>
<meta:initial-creator>Dmitri Popov</meta:initial-creator>
<meta:creation-date>2006-11-30T11:47:40</meta:creation-date>
<dc:creator>Dmitri Popov</dc:creator>
<dc:date>2006-11-30T11:58:51</dc:date>
<dc:language>en</dc:language>
```

## Co w plikach piszczy

Każdy z plików zapisanych wewnątrz dokumentu ODF ma określone przeznaczenie. Główny plik to *content.xml*, który przechowuje faktyczną treść dokumentu. Wystarczy otworzyć *content.xml* w edytorze tekstowym, by przekonać się, że użyta w nim składnia XML przypomina stary dobry HTML. Nie trzeba być ekspertem od języka XML, żeby domyślić się znaczenia podstawowych znaczników. Ciąg „Lorem ipsum” z Listing 1 jest na przykład oznaczony jako nagłówek pierwszego rzędu (heading 1). Ciąg „Proin velit” zostanie natomiast w sformatowanym tekście wyfłuszczony.

Tak jak tradycyjny plik HTML, *content.xml* zawiera właściwy tekst oraz instrukcje formatujące. Polecenia formatujące przypisują do fragmentów tekstu określony styl i krój pisma. Bardziej złożone dokumenty zawierają dodatkowe znaczniki, na przykład określające położenie ramek tekstowych i obrazów.

Pozostałe pliki w archiwum to albo treści uzupełniające (takie jak osadzone w dokumencie obrazy), albo inne dodatkowe informacje. Liczba plików waha się w zależności od rodzaju dokumentu i aplikacji, która go zapisała. Kilka z nich zasługuje na szczególną uwagę.

Plik *manifest.xml* zawiera spis wszystkich plików w dokumencie (Listing 2). Znajduje się w katalogu *META-INF*.

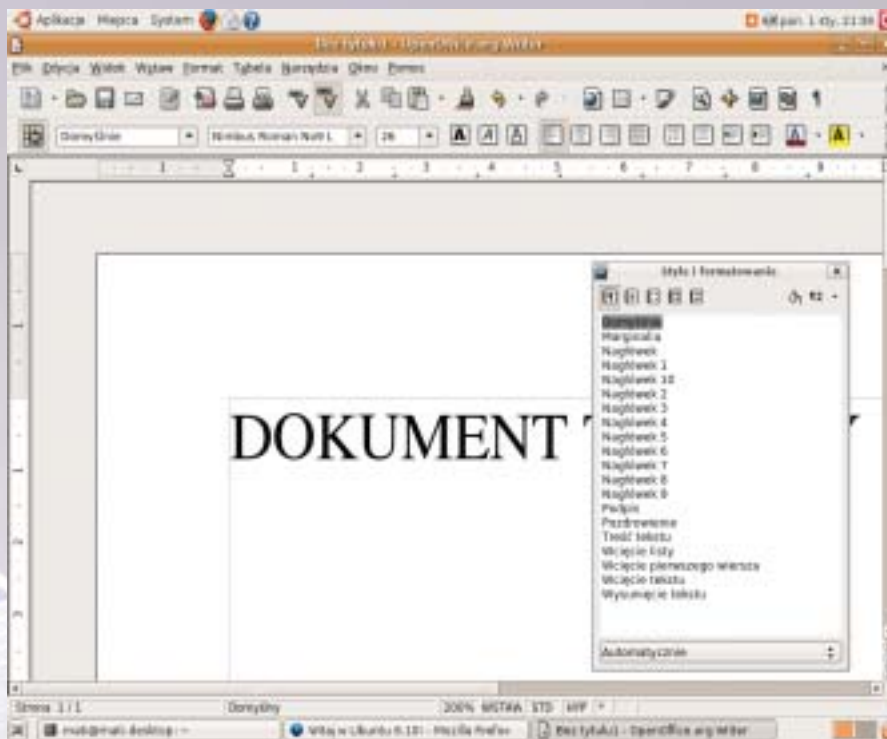
Plik *meta.xml* zawiera metadane - między innymi nazwę aplikacji zapisującej dokument, nazwisko autora, datę utworzenia, język i liczbę słów (Listing 3). Uważni Czytelnicy zauważą, że podobne informacje wyświetlają się w oknie właściwości dokumentu (Document Properties) programu OpenOffice.org (Rysunek 1).

Następny plik, *settings.xml*, opisuje różnorodne ustawienia związane z samym dokumentem. Należą do nich takie parametry jak: Field Auto Update (automatyczna aktualizacja pola), Save Version on Close (aktualizacja numeru wersji przy zamknięciu), Printer Name (nazwa drukarki), itp. Specyfikacja formatu podaje, że *settings.xml* jest przeznaczony na te parametry, których używa aplikacja, w której dokument został utworzony, ale niekoniecznie inne aplikacje, w których może być otwierany.

Większość współczesnych edytorów tekstu pozwala przypisać danemu akapitowi określony styl (Rysunek 2). Plik *styles.xml* definiuje style związane z określonym dokumentem. Style umożliwiają oddzielenie for-

Tabela 1: Typy plików OpenDocument

Rozszerzenie	Opis	Typ MIME
odt	tekst	application/vnd.oasis.opendocument.text
ott	tekst - szablon	application/vnd.oasis.opendocument.text-template
oth	tekst - szablon HTML	application/vnd.oasis.opendocument.text-web
odm	dokument główny	application/vnd.oasis.opendocument.text-master
odg	grafika wektorowa	application/vnd.oasis.opendocument.graphics
otg	grafika wektorowa - szablon	application/vnd.oasis.opendocument.graphics-template
odp	prezentacja	application/vnd.oasis.opendocument.presentation
otp	prezentacja - szablon	application/vnd.oasis.opendocument.presentation-template
ods	arkusz kalkulacyjny	application/vnd.oasis.opendocument.spreadsheet
ots	arkusz kalkulacyjny - szablon	application/vnd.oasis.opendocument.spreadsheet-template
odc	wykres	application/vnd.oasis.opendocument.chart
otc	wykres - szablon	application/vnd.oasis.opendocument.chart-template
odi	grafika	application/vnd.oasis.opendocument.image
oti	grafika - szablon	application/vnd.oasis.opendocument.image-template
odf	wzory matematyczne	application/vnd.oasis.opendocument.formula
otf	wzory matematyczne - szablon	application/vnd.oasis.opendocument.formula-template



Rysunek 2: Większość edytorów tekstu pozwala użytkownikowi wykorzystywać do formatowania gotowe style.

matowania od treści, dzięki czemu główny plik *content.xml* jest prostszy i łatwiej nim zarządzać.

Plik mimetype opisuje – jak łatwo zgadnąć – typ MIME, czyli rodzaj dokumentu. Najpopularniejsze typy dokumentów ODF prezentuje Tabela 1. W omawianym przykładzie – dokumencie tekstowym – typ MIME to *application/vnd.oasis.opendocument.text*.

## Automatyzacja

Język XML z założenia umożliwia przetwarzanie automatyczne a prostota i czytelność formatu ODF pozwalają programistom łatwo tworzyć narzędzia do jego obsługi. Liczba możliwych ustawień i opcji ODF jest oczywiście ogromna. Z wyjątkiem wąskiej grupy aplikacji, nie trzeba przetwarzać wszystkich możliwych opcji formatowania i konfiguracyjnych.

## A co to jest otwarty format?

Nie istnieje wprawdzie formalna definicja „otwartego formatu”, nietrudno jednak wymienić podstawowe warunki, jakie musi on spełniać.

Po pierwsze: otwarta, publicznie i jawnie rozwijana specyfikacja. ODF, jak każdy otwarty format, powstał w rezultacie licznych dyskusji i długotrwałych prac wielu grup ludzi. Za jego kształt odpowiadają firmy i instytucje będące członkami OASIS, organizacji koordynującej rozwój i promowanie otwartych standardów (w szczególności samego ODF). Członkowie OASIS bezpośrednio uczestniczą w tworzeniu specyfikacji ODF, każdy może jednak zaangażować się w rozwój formatu i uzyskać dostęp do interesujących go informacji. OASIS to nie grupa trzymająca władzę - każdy może się do niej wprosić, a już na pewno śledzić postępy prac. Aby włączyć się w prace o OASIS, warto zajrzeć najpierw na strony OASIS OpenDocument Technical Committee [7], OASIS ODF Adoption Committee [8] oraz OpenDocumentXML.org [9].

Warunek drugi: naprawdę otwarty standard musi być wolny od ograniczeń patentowych i licencyjnych. To pozornie oczywisty warunek, ale absolutnie nie wolno go pomijać. Zwalnia programistów od konieczności uzyskania licencji lub wnoszenia opłat autorskich za prawo wykorzystywania. Użytkownik nie musi się z kolei martwić, czy zapisując dokumenty w tym formacie nie narusza czyichś patentów lub nie łamie postanowień licencji. Niektórzy producenci i niektóre organizacje mają tendencje do używania słowa „otwarty” w odniesieniu do standardów, których inni producenci i inne organizacje raczej za takie nie uznają. Złożony problem ograniczeń patentowych i licencyjnych to temat nieustannej dyskusji twórców oprogramowania.

Dalej, otwarty format nie może bazować na składnikach, które są zamknięte i dostarczane tylko przez jednego producenta. ODF nie polega na żadnych zamkniętych technologiach i nie zezwala korporacjom na dowolne dodawanie funkcji. Z tego względu nie grożą mu takie praktyki, jakie niegdyś Microsoft zastosował względem Javy. Dodanie autorskich rozszerzeń sprawiło wówczas, że Java Microsoftu stała się niekompatybilna z Javą reszty świata.

Jako otwarty format ODF zapewnia też programistom łatwość implementacji, oraz łatwość współdziałania dowolnych posługujących się nim programów. Ponieważ jego struktura jest jawna i dobrze udokumentowana, deweloperzy mogą dość łatwo uzupełniać istniejące programy o obsługę tego formatu oraz tworzyć nowe programy, które z niego korzystają. Listy programów obsługujących obecnie ODF można znaleźć w Wikipedii [10], i na stronie OpenDocument Fellowship [11] - pod hasłem Supported Software.

Z nadejściem Office 2007, Microsoft ogłosił wprowadzenie nowego, opartego na XML formacie dokumentu. Podstawowe różnice, z punktu widzenia Open Source, między formatem OpenDocument a formatem Microsoftu, przedstawia artykuł „Format Comparison Between ODF and MS XML” [12]. Krótkie streszczenie tego artykułu można znaleźć na stronach Wikipedii [13].

Aby lepiej zrozumieć różne aspekty definiowania otwartych standardów, warto także posłuchać wywiadu z Danese Cooper na stronach Open Source Conversations [14].

Aby bezboleśnie przygotować program zapisujący pliki w formacie ODF, najlepiej zacząć od zapisania wzorcowego dokumentu, zawierającego wszystkie wymagane cechy i elementy formatowania, a następnie rozpakować powstały plik ODF (w sposób omówiony na początku artykułu) i przeanalizować składnię XML. Eksperymentując z poszczególnymi ustawieniami XML i przeladując dokument ODF, można od razu przekonać się, jak wprowadzane zmiany wpływają na postać końcową dokumentu. A potem – jak wdzięcznie ujął to J. David Eisenberg w OASIS OpenDocument Essentials [2]: „Gdy przetestujesz działanie danej funkcji, skopiuj kod XML z dokumentu do swojego programu, czyli – po prostu ściągaj. Wiem, co mówię: sam tak robiłem pisząc tę książkę”.

O wiele łatwiej odratować lub naprawić plik XML niż plik binarny. Nawet jeśli określona aplikacja nie zdoła samodzielnie otworzyć uszkodzonego dokumentu, odzyskanie z niego danych nie sprawi kłopotu. W przeciwieństwie do plików binarnych, w plikach XML treść jest przecież zapisywana w postaci zrozumiałej dla istoty ludzkiej. Dobrym przykładem prostego, ale efektywnego narzędzia do konwersji i odzyskiwania plików tekstowych ODF jest skrypt w języku Python odt2txt [3]. Skrypt konwertuje dokumenty OpenOffice.org Writer na pliki tekstowe, sprrowadzając formatowanie do postaci składni Markdown.

### Podsumowanie

Choć format ODF stanowi interesującą otwartą alternatywę do formatów zamknię-

tych, ma i kilka słabych stron. Kontrowersje budzi na przykład kwestia wydajności formatu przy przetwarzaniu dużych ilości danych (por. [4] i [5]).

Im większą rolę w społeczeństwie odgrywają dokumenty składowane i wymieniane w postaci elektronicznej, tym bardziej istotny staje się problem współdziałania programów i zakłóceń przepływu informacji spowodowanych stosowaniem zamkniętych formatów. W przypadku ODF takie problemy nie występują. Prędzej lub później decydenci polityczni i użytkownicy biznesowi przekonają się, że warto przechowywać dane w otwartych formatach – takich jak ODF. ■

## INFO

- [1] OASIS: <http://www.oasis-open.org/>
- [2] OASIS OpenDocument Essentials, by J. David Eisenberg: <http://books.evc-cit.info/>
- [3] odt2txt: <http://www.freewisdom.org/projects/python-markdown/odt2txt.php>
- [4] Rozważania na temat ODF: <http://blogs.zdnet.com/Ou/?p=101>
- [5] Kolejne uwagi na temat ODF: <http://blogs.zdnet.com/Ou/?p=119>
- [6] „Why OpenDocument Won (and Microsoft Open XML Didn't)": <http://www.dwheller.com/essays/why-opendocument-won.html>
- [7] OpenDocument Technical Committee: [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=office](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office)
- [8] OASIS ODF Adoption Committee: [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=odf-adoption](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=odf-adoption)
- [9] OpenDocumentXML.org: <http://opendocument.xml.org/>
- [10] Informacje z Wikipedii o programach obsługujących OpenDocument: [http://en.wikipedia.org/wiki/OpenDocument\\_software](http://en.wikipedia.org/wiki/OpenDocument_software)
- [11] Programy obsługujące OpenDocument: <http://opendocumentfellowship.org/applications>
- [12] „Format comparison between ODF and MS XML": <http://www.groklaw.net/articlebasic.php?story=20051125144611543>
- [13] Wikipedia - porównanie ODF i MS XML: [http://en.wikipedia.org/wiki/Comparison\\_of\\_OpenDocument\\_and\\_Microsoft\\_Office\\_Open\\_XML\\_formats](http://en.wikipedia.org/wiki/Comparison_of_OpenDocument_and_Microsoft_Office_Open_XML_formats)
- [14] Wypowiedź Danese Cooper o otwartych standardach: <http://osc.gigavox.com/shows/detail1222.html>