

## Wykład 1: Tekstowe dokumenty elektroniczne

### 1 Formaty dokumentów

Publikując dla uczniów/studentów/słuchaczy materiały w sieci musimy rozważyć najpierw **format plików**<sup>1</sup>, którego użyjemy w naszej publikacji. Dotyczy to oczywiście nie tylko publikacji tekstowych, więc do zagadnienia formatów plików będziemy wracać przy każdym rodzaju multimediiów.

Dokonując wyboru formatu musimy wziąć pod uwagę następujące czynniki:

**Zawartość.** Oprócz tekstu, w dokumencie będą znajdowały się prawdopodobnie inne media (choćby ilustracje) — trzeba wybrać format odpowiedni do ich przechowywania.

**Rozmiar.** Skoro nasz materiał ma być dostępny w formie elektronicznej, to nie bez znaczenia jest wielkość pliku. Ta sama zawartość w różnych formatach będzie generowała pliki różnych rozmiarów — a różnice mogą być nawet kilkunastokrotne!

**Otwartość.** Niektóre formaty są zamknięte — czyli opracowane i w pełni kontrolowane przez jedną firmę, która zachowuje monopol na narzędzia poprawnie wyświetlające dany format. Na publikowanie w formatach zamkniętych powinniśmy z czystym sumieniem decydować się jedynie wtedy, gdy mamy pewność, że wszyscy korzystający z naszych materiałów mają swobodny dostęp do odpowiedniego oprogramowania komercyjnego.

**Dostępność oprogramowania.** Wiąże się w jakiś sposób z otwartością formatu. Jeżeli nie mamy pewności, że użytkownik może swobodnie przeglądać publikację, powinniśmy wska-

<sup>1</sup>[http://pl.wikipedia.org/wiki/Format\\_pliku](http://pl.wikipedia.org/wiki/Format_pliku)

zać odpowiednie oprogramowanie darmowe (przynajmniej jako alternatywę dla komercyjnego). W przypadku formatów otwartych nie ma z tym zwykle problemu.

**Zmienność.** Zwykle pożądanym jest, by nasz dokument miał formę dokładnie taką, jaką my mu nadaliśmy. Niestety, wiele formatów ma to do siebie, że dokumenty w nich zapisane różnie prezentują się na różnych komputerach lub pod różnym oprogramowaniem.

**Cel.** Co studenci mają zrobić z naszym plikiem? Inny format powinniśmy wybrać, gdy chcemy tylko, by studenci zapoznali się z materiałem, a inny, gdy oczekujemy od nich jakiejś pracy (uzupełnienia, dopisków, przeredagowania itp.).

**Redagowalność.** Wiąże się nieco z poprzednim punktem — ale zwykle zależy nam na tym, by ci, którzy otrzymują od nas materiały nie mogli ich w prosty sposób pozmienić (i, na przykład, wykorzystać jako swoich). Niektóre formaty są wprost przewidziane do edycji swojej zawartości, inne są bardzo niewygodne w redakcji zawartości, a jeszcze inne oferują wręcz bardzo zaawansowane zabezpieczenia przed zmianami.

## 1.1 Formaty zalecane

### 1.1.1 PDF

*Portable Document Format*<sup>2</sup> jest idealnym formatem do tekstowych publikacji internetowych<sup>3</sup>. Przede wszystkim pozwala na prezentację tekstu z zachowaniem znaków narodowych, a oprócz tekstu można zamieszczać w dokumentach tego formatu także grafikę wektorową i rastrową, interaktywną grafikę trójwymiarową i wiele innych typów zawartości. Pliki nie są zbyt duże (szczególnie, jeśli nie mają być przeznaczone do druku — wtedy grafika może być w słabszej rozdzielczości) ze względu na wewnętrzną kompresję danych. Standard jest otwarty, więc istnieje wiele programów do jego przeglądania (w tym dobrze znany *Adobe Reader*<sup>4</sup>), jak i do

<sup>2</sup>[http://pl.wikipedia.org/wiki/Portable\\_Document\\_Format](http://pl.wikipedia.org/wiki/Portable_Document_Format)

<sup>3</sup>PDF nadaje się też niezłe jako format prezentacji ekranowych — ale o tym na wykładzie o prezentacjach ekranowych właśnie.

<sup>4</sup><http://www.adobe.com/products/reader/>

tworzenia plików tego formatu (jak na przykład *QPrinter*<sup>5</sup> pod MS Windows; wiele innych programów potrafi tworzyć pliki PDF z własnych dokumentów, jak *OpenOffice.org*<sup>6</sup>).

Pliki PDF są pojedynczymi dokumentami, więc nadają się do przekazywania zwartych całości — jak książki, wykłady itp., które mają być i tak ściągnięte w całości na komputer odbiorcy i tam czytane (po ewentualnym wydrukowaniu). Poza tym format PDF zapewnia **dokładnie taki sam wygląd** na komputerze oglądającego (i ewentualnym wydruku), jak na komputerze autora. Nie tylko dotyczy to układu tekstu i grafiki, ale także użytych czcionek, które mogą zostać w razie potrzeby osadzone w dokumencie<sup>7</sup>. Ponadto dokument w tym formacie nie jest przeznaczony do zmian, więc studenci nie mogą w prosty sposób przeredagowywać go (choć nie jest to całkiem niemożliwe).

### 1.1.2 HTML

Alternatywą dla formatu PDF jest prezentacja treści w formacie *HTML*<sup>8</sup> (czy też *XHTML*<sup>9</sup>). Jest to uniwersalny język opisu stron internetowych. Nie powstał on jednak po to, by zapewnić im jednolity wygląd i każda przeglądarka prezentuje dokumenty Internetowe nieco inaczej. Dokumenty HTML nie zawierają też opisu czcionek, więc jeżeli odbiorca nie ma zainstalowanego odpowiedniego kroju, przeglądarka użyje innego, czasem bardzo różnego od zaplanowanego przez twórcę. Co więcej, jej użytkownik może zwykle pozmieniać wiele parametrów wyświetlania, tak, by ustawić swoje czcionki, kolory, rozmiary... Poza tym dokument będzie inaczej wyglądał przy zmianie rozmiaru okna, w którym jest wyświetlany, bo przeglądarka za każdym razem dostosowuje układ i łamanie tekstu do dostępnego obszaru.

Jednakże HTML i XHTML są standardami i w praktyce każdy komputer podłączony do sieci posiada oprogramowanie do przeglądania tego rodzaju plików. Ponadto materiały przygotowane w tych formatach nie stanowią pojedynczych, zwartych plików (zwykle podzielone są na podstrony, wyodrębnione są też grafiki i inne materiały multimedialne), więc komputer odbiorcy nie potrzebuje ściągać całości materiału, lecz tylko te fragmenty, które jego użytkownik

<sup>5</sup><http://www.qprint.com.pl/index.pl/qprinter>

<sup>6</sup><http://pl.wikipedia.org/wiki/OpenOffice.org>

<sup>7</sup>To znaczy, że opis wyglądu czcionek załączany jest w samym pliku.

<sup>8</sup><http://pl.wikipedia.org/wiki/HTML>

<sup>9</sup><http://pl.wikipedia.org/wiki/XHTML>

w danej chwili chce przejrzeć. W sposób naturalny zmniejsza to obciążenie sieci i przyspiesza dostęp do poszukiwanego materiału — pod warunkiem, że chcemy zapewnić doń szybki dostęp i możliwość przeglądania on-line, bez konieczności ściągania całości (jak w przypadku słowników, encyklopedii, leksykonów. . .).

### 1.1.3 OpenDocument

*OASIS Open Document Format for Office Applications*<sup>10</sup> (bo tak brzmi pełna oficjalna nazwa tego formatu) to otwarty standard (*de iure*, zatwierdzony przez ISO<sup>11</sup>) dokumentów biurowych określający format dla dokumentów tekstowych, arkuszy kalkulacyjnych, prezentacji ekranowych, grafiki, wykresów, wzorów matematycznych. Oparty jest na języku XML<sup>12</sup> i formacie plików wprowadzonym w pakiecie biurowym *OpenOffice.org 2.0*. Ponieważ standard jest względnie prosty, istnieje wiele aplikacji obsługujących go (lepiej lub gorzej), w tym przede wszystkim OpenOffice.org, dostępny bezpłatnie i z otwartym kodem źródłowym.

Jeżeli potrzebujemy opublikować plik, w którym studenci mają dokonywać jakichś zmian (jak formularz do uzupełnienia, arkusz kalkulacyjny do obliczeń. . .), to polecanym formatem jest właśnie OpenDocument. Pliki w nim zapisane są względnie niewielkie, a raczej każdy student może legalnie ściągnąć i zainstalować na swoim komputerze darmowe (ale w pełni funkcjonalne) oprogramowanie do jego obsługi.

## 1.2 Inne formaty

W tej sekcji omówimy kilka popularnych formatów, które jednak nie są w normalnych sytuacjach zalecane do publikacji w nauczaniu zdalnym (ani w ogóle w Internecie). Można jednak wyobrazić sobie sytuacje, gdy możemy (lub nawet musimy) ich używać — choćby format PPT przy tworzeniu kursu „Podstawy grafiki prezentacyjnej w programie MS PowerPoint” . . .

<sup>10</sup><http://pl.wikipedia.org/wiki/OpenDocument>

<sup>11</sup><http://pl.wikipedia.org/wiki/ISO>

<sup>12</sup><http://pl.wikipedia.org/wiki/XML>

### 1.2.1 RTF

*Ritch Text Format*<sup>13</sup> to otwarty format opracowany i szczegółowo wyspecyfikowany przez Microsoft dla wymiany dokumentów tekstowych pomiędzy różnym oprogramowaniem. Jest dobrze zdomowionym w oprogramowaniu do obróbki tekstu sposobem zapisu plików, obsługiwanym przez wiele programów tego rodzaju. Pliki tego formatu nie są też duże. Dlatego też, format RTF lepiej nadaje się do publikacji od formatu DOC (patrz niżej). Niestety nie każde formatowanie może być zapisane w plikach tego rodzaju, a poza tym pliki te przeznaczone są do edycji, więc narażone na niepożądane zmiany ze strony studentów — ma więc RTF dość ograniczone zastosowanie.

### 1.2.2 TXT

Format ten to czysty tekst bez żadnego formatowania. Nadaje się on przed wszystkim do publikowania kodów źródłowych programów. Ograniczenie w jego zastosowaniu wynika głównie z różnic w *kodowaniu poszczególnych znaków*<sup>14</sup> (przede wszystkim *znaku końca linii*<sup>15</sup>) w różnych systemach operacyjnych czy zastosowaniach.

### 1.2.3 Dawne formaty MS Office

Z powodu ogromnej popularności pakietu *Microsoft Office*<sup>16</sup> bardzo rozpowszechnione są formaty DOC, XLS, PPT — domyślnie używane przez wszystkie kolejne (bez najnowszej) wersje tego oprogramowania. Mają one jednak wiele cech, które (w przekonaniu autora i wielu użytkowników Internetu) dyskwalifikują je jako formaty szerokiej (internetowej) wymiany danych. Dlaczego? Wśród tych wad wymienić należy:

**Zamkniętość.** Specyfikacje tych formatów nie są oficjalnie dostępne, więc nie ma możliwości dobrej (dokładnej) implementacji oprogramowania do obsługi tego rodzaju plików, a równocześnie niezależnej od firmy Microsoft. Owszem, istnieje oprogramowanie radzące sobie

<sup>13</sup><http://pl.wikipedia.org/wiki/RTF>

<sup>14</sup>[http://pl.wikipedia.org/wiki/Zestaw\\_znakow](http://pl.wikipedia.org/wiki/Zestaw_znakow)

<sup>15</sup><http://pl.wikipedia.org/wiki/End-of-line>

<sup>16</sup>[http://pl.wikipedia.org/wiki/Microsoft\\_Office](http://pl.wikipedia.org/wiki/Microsoft_Office)

(lepiej lub gorzej) z tymi plikami, ale albo nie radzi sobie z nimi idealnie (choć w wielu wypadkach bardzo dobrze — jak OpenOffice.org), albo jest dość drogie.

**Przywiązanie do jednego, kosztownego oprogramowania.** Wada ta wynika wprost z poprzedniej. Oczywiście, jeżeli mamy pewność, że wszyscy nasi studenci mają swobodny i legalny dostęp do tego oprogramowania (lub możemy ich zmusić do zakupu takiego) — nie jest to przeszkodą. Pamiętać tylko należy, że zwiększa to realny koszt prowadzonego przez nas kursu czy też szkolenia. . .

**Rozmiar.** Pliki omawianych tu formatów zwykle są dość duże — szczególnie dlatego, że zawierają mnóstwo dodatkowych danych, które i tak zwykle nie są potrzebne odbiorcy.

**Zmienność.** Nigdy nie możemy być pewni, że plik starannie sformatowany na naszym komputerze będzie wyglądał dokładnie tak samo na komputerze studenta. Właściwie, możemy być wręcz pewni, że mniejsze lub większe różnice w układzie się pojawią!

**Redagowalność.** W oczywisty sposób pliki DOC, XLS, PPT i podobne mogą być bez przeszkód zmieniane (bo te formaty służą do edycji) przez studentów — wbrew naszym intencjom.

#### 1.2.4 Nowe formaty MS Office (OOXML)

*Office Open XML*<sup>17</sup> jest zestawem formatów (DOCX, XLSX, PPTX i inne) nowo opracowanym przez firmę Microsoft jako odpowiedź na standard OpenDocument. Pomimo otwartości formatu i uznania za standard przez ISO, w tej chwili OOXML w roli uniwersalnego sposobu zapisu plików biurowych nie sprawdza się. Dzieje się tak dlatego, że format jest względnie nowy, więc nie zdążyły jeszcze powstać aplikacje obsługujące go (poza MS Office 2007 i OpenOffice.org w wersji 3.0), a złożoność formatu (szczególnie w porównaniu z OpenDocument) także nie sprzyja adaptowaniu go w oprogramowaniu.

Z tych względów, na chwilę obecną, OOXML nie stanowi dobrego wyboru dla publikacji Internetowych, w tym edukacyjnych i jak na razie należałoby unikać używania tych formatów w edukacji.

<sup>17</sup>[http://pl.wikipedia.org/wiki/Office\\_Open\\_XML](http://pl.wikipedia.org/wiki/Office_Open_XML)