

Zadania domowe – Perl – Przetwarzanie XML

Termin złożenia zadania domowego to 21.05.2008. Maksymalna liczba punktów za zadania XML1 do XML6 wynosi 20 punktów. Punkty za zadania dodatkowe zostaną doliczone niezależnie. Łącznie można zdobyć 34 pkt.

Zadanie XML1

Proszę rozszerzyć program z zestawu „Przetwarzanie tekstów 2” w następujący sposób:

- a) Zamiast wersji tekstowej program powinien generować informacje zapisane w XML.
- b) Opisać proszę wybrany schemat XML i uzasadnić wybór znaczników, zadbać proszę o czytelność dokumentu.
- c) Napisać odpowiedni DTD (wewnętrzny) dla wykorzystanego fragmentu XML. Upewnić się, że tak opisany plik XML waliduje z podanym DTD (przy wykorzystaniu wewnętrznego DTD i podanego walidatora).

Punktacja: 1 pkt. za każdy podpunkt.

Zadanie XML2

Proszę rozszerzyć powyższy program tak, aby wczytywał kilka plików podanych w wierszu poleceń (np. za pomocą tablicy @ARGV). Informacje wyekstrahowane z plików HTML mają być zapisane do jednego, wspólnego pliku XML. Rozszerzyć w tym celu nasz podjęzyk XML oraz DTD. Upewnić się, że nowy plik waliduje.

Punktacja: 2 pkt. (+2 pkt. jeśli program będzie wczytywał wszystkie pliki *.html w danym katalogu)

Zadanie XML3

Proszę dokończyć zadanie z zajęć praktycznych dotyczące tworzenia podjęzyka XML opisującego drzewa takie jak w pliku <http://www.logic.amu.edu.pl/images/5/52/Syntaxtree.png>.

- a) Stworzyć plik XML, który opisuje dane drzewo.
- b) Stworzyć odpowiedni DTD (wewnętrzny) i upewnić się, że dokument waliduje.
- c) Opisać w taki sam sposób zdanie „There are lots of trees in a forest” i upewnić się, że nowy plik waliduje z poprzednim DTD.
- d) Rozszerzyć podjęzyk XML oraz DTD tak, aby można było zapisywać wiele drzew w jednym dokumencie XML.

Punktacja: 2 pkt. za każdy podpunkt.

Zadanie XML4

Napisać za pomocą modułu XML::Simple program, który wczytuje dokument XML generowany przez program z zadania XML2. Program ma wyświetlić na wyjściu standardowym po kolei nazwisko autora, tytuł artykułu oraz datę dla każdego artykułu.

Punktacja: 4 pkt.

Zadanie XML5

Załóżmy, że plik XML z zadania XML2 lub XML4 zawiera kilka roczników wiadomości z Gazety Wyborczej. Mamy więc pewien korpus artykułów prasowych o pokaźnych rozmiarach. Napisać program podobny do zadania XML4, który nie wczytuje całego pliku XML do pamięci operacyjnej komputera:

- a) Bez wykorzystania modułu XML::Simple. Skoro jesteśmy zainteresowani tylko nazwiskiem, tytułem oraz datą, to możemy wykorzystać np. wyrażenia regularne.
- b) Wykorzystując XML::Simple. Wskazówka: XML ma taką własność, że każdy podelement można traktować jako element główny innego mniejszego dokumentu XML.

Punktacja: 3 pkt. za każdy podpunkt.

Zadanie XML6

Napisać nową wersję programu z PERL2.3, PERL3.3 i PERL4.7 do odmiany niemieckich przymiotników – regularnych oraz nieregularnych. Zapisać wszelkie informacje lingwistyczne w zewnętrznym pliku XML-owym. Informacjami lingwistycznymi nazywamy informacje o sufiksach, infiksach oraz formy podstawowe i stopniowane przymiotników nieregularnych. Program ma pobierać wszystkie informacje z pliku XML z pomocą XML::Simple. Zaproponować podjęzyk XML służący do opisu takich danych.

Punktacja: 5 pkt.

Zadanie dodatkowe XML7

Napisać program podobny do XML5. W punkcie a) zamiast wyrażen regularnych wykorzystać moduł XML::SAX::Expat; W punkcie b) wykorzystać moduł XML::Twig.

Punkty: 7 pkt. za każdy podpunkt niezależnie od maksymalnej liczby punktów.