Konwersja ISO->UTF i na odwr□łt

Autor: Jokris
17.01.2008.

O konwersji system□w kodowania znak□w pisaÂłem wielokrotnie na Forum Jokris.info, np. TUTAJ. Jeden z uÂżytkownik□w mojego forum napisaÂł: "Probowalem skonwertowa□Ś plik polish.php do utf-8, ale nie przyniosÂło to efektu. Krzaczki byÂły zast□□powane przez troch□□ inne krzaczki . Program do konwersji (Edit pad pro) twierdzi, Âże plik polish.php jest kodowany (original encoding) w Windows 1252: Western European." ChociaÂż kaÂżdy ma sw□j spos□b na brak polskich znak□w, podam wam opis metody, kt□rÂą preferuj□□ja. I przynosi zawsze pozytywny efekt ko□cowy. Temat ten jest jak najbardziej aktualny, bo ostatnio, sÂłusznie czy nie sÂłusznie, cz□□Âś□Ś uÂżytkownik□w przestawiÂła si□□na kodowanie UTF-8 dla Joomla. Na poczÂątek waÂżna informacja: ÂŻeby przekonwertowa□Ś (przekodowa□Ś) tekst z ISO-8859-2 na UTF-8 naleÂży najpierw dokona□Ś konwersji tekstu z ISO-8859-2 na Windows-1250. Teraz troch□□teorii na temat system□w kodowa□ą znak□w . Windows-1250 jest to strona kodowa kt□rej uÂżywacie na co dzie□ą np. w Notatniku systemowym, Wordzie, czy chociaÂżby wprowadzajÂąc tekst z klawiatury do postu na Forum. Czyli z polskimi znakami diakrytycznymi, jak np "ÂŚ, Âś, Âą, Âź" i.t.d. Bo musicie wiedzie□Ś, Âże ISO-8859-2 jest kojarzone ze standardem polskich znak□w, ale tak w rzeczywistoÂści to wszystkie litery z ogonkami zastÂąpione sÂą innymi znakami, kt□re majÂą z j□□zykiem polskim tyle wsp□lnego, co nic. Dopiero przeglÂądarka internetowa rozkodowuje tekst (po to mamy w menu przeglÂądarki list□□ z kodowaniem znak□w), i przedstawia go juÂż z polskimi ogonkami. Dlatego nazywa si□□to "SYSTEM KODOWANIA ZNAK□□W". ISO-8859-2, czy teÂż UTF-8 sÂą to systemy kodowania, kt□re zamieniajÂą polskie ogonki (w przypadku ISO-8859-2) i r□wnieÂż polskie nietypowe znaki, jak np. "Âł, Âż" (w przypadku UTF-8) na ich odpowiedniki zakodowane (po prostu zamienione) w okreÂślonym w w/w systemach kodowania. I zamiast literki Âą b□□dziesz miaÂł znak ą. Troch□□teorii a teraz do konkret□w. JeÂśli te "konkrety" interesujÂą Was, kliknijcie w poniÂższy odnoÂśniczek...

- Do przekodowywania pomi□□dzy systemami kodowania znak□w uÂżywaj jedynego, sprawdzonego programu o nazwie GÂżegÂż□Âłka XP. GÂżegÂż□ÂłkaÂ to konwerter standard□w kodowania polskich znak□w diakrytycznych (zwanych potocznie ogonkami). Opr□cz ogonk□w obsÂługuje r□wnieÂż wi□□kszoÂś□Ś standard□w kodowania stosowanych na caÂłym Âświecie.
- Jak robi□Ś konwersj□□?. Po prostu wystarczy otworzy□Ś np. taki plik "polish.php" za pomocÂą drugiego, jedynego i sprawdzonego programu (a co, jak za komuny) jakim jest Notatnik SP PL. Ja w tym edytorze przetÂłumaczyÂłem wiele plik□w jak i teÂż edytuj□□kod PHP lub HTML. Nie znam lepszego. (oczywiÂście to tylko moja opinia, nie zawsze jedyna i sÂłuszna).
- Po otwarciu pliku z kodowaniem znak□w ISO-8859-2, wystarczy zaznaczy□Ś w menu "Konwersja" => kodowanie Windows-1250". I juÂż zamiast dziwnych znak□w b□□dziesz miaÂł polskie literki. A musicie to zrobi□Ś dlatego, aby GÂżegÂż□ÂłkaÂ XP poprawnie rozpoznaÂł Âźr□dÂło, i system znak□w jakie ma przekonwertowa□Ś.

{mosimage cw=300 popup=1 popupTyp=dhtml}

- Po tej operacji (zawsze r□bcie kopie oryginalnych plik□w) uruchamiasz program GÂżegÂż□ÂłkaÂ XP i pojawi Ci si□□jego okienko. Wrzucasz do niego metodÂą "przeciÂągnij - upuÂś□Ś" plik "polish.php", ale oczywiÂście ten w zestawie znak□w Windows-1250. Na dole masz 2 pola typu "lista", rozwijalne. W lewym polu wybierasz zestaw znak□w "Windows 1250(Europa ÂŚrodkowa)", natomiast w prawym polu, kt□ry jest docelowym, wybierasz interesujÂący Ci□□zastaw znak□w, czyli "Unicode UTF-8". Naciskasz tylko "Start". Potwierdzasz, Âże chcesz dokona□Ś konwersji, i w miejscu pliku "polish.php" zakodowanym w standardzie Windows 1250 pojawi si□□"polish.php" w kodowaniu UTF-8. Program automatycznie wykonuje kopi□□oryginalnego pliku "polish.php". I po wszystkim.

Proste?. No pewnie teraz tak. Musicie pami□□ta□Ś, Âże po przekodowaniu pliku przykÂładowego "polish.php" z ISO-8859-2 na UTF-8, w zawsze sÂłusznym Notatniku SP PL nie zobaczycie polskich znak□w, tylko bardzo dziwne "krzaczki", kt□re to w/w edytor nie potrafi wyÂświetli□Ś. ZresztÂą jak i wiele innych edytor□w. Sprawdzicie poprawnoÂś□Ś kodowania w Notatniku systemowym. Tam powinniÂście mie□Ś polskie znaki. JeÂśli sÂą, oznacza to, Âże konwersja udaÂła si□□. Notatnik systemowy obsÂługuje kodowanie UNICODE. Szkoda Âże nie majÂą tej cechy edytory, chociaÂżby dost□□pne na mojej stronie.

Jokris (Krzysiek Stachyra). 18 stycznia 2008r.