

Unicode in Homesite 5

Es geht - nicht wirklich

Pseudo-Unicode in Homesite

Die Unicode-Unterstützung von Homesite ist zum Bedauern der Nutzer dieses hervorragenden Quelltext-Editors verwirrend und im Großen und Ganzen mangelhaft.

Es ist möglich, im Homesite-Editor mit den unter Options/Settings/Editor/Charset angegebenen Zeichenkodierungen bzw. Schriftzeichen Texte einzugeben, vorausgesetzt, in Windows ist die Mehrsprachenunterstützung aktiviert und auch das entsprechende Eingabeschema aktiviert. Leider hat das Ganze mehr Spaßfunktion und macht nicht wirklich Sinn.

Achtung: Die Standard-Einstellung bei der Homesite-Charset Einstellung ist "default"; die ausgewählte Zeichenkodierung bezieht sich nur auf die Darstellung der Zeichen während ihrer Eingabe im Homesite-Editor selbst.

So kann man beispielsweise im Homesite-Editor arabisch schreiben:

1. Unter Options/Settings/Editor/Charset die arabische Zeichenkodierung wählen (dient wie gesagt nur der korrekten Anzeige im Homesite-Editor)
2. In Windows das Gebietsschema AR auswählen, z.B. über die Eingabegebietsschemaleiste
3. Losschreiben

Die Eingabe wird im Homesite-Editor korrekt angezeigt und man kann den eingegebenen Text auch problemlos in Notepad hineinkopieren und umgekehrt.

Doch diese Pseudo-Unicode-Eingabe bringt nichts, denn jedes so eingegebene Zeichen wird vom Homesite-Editor falsch kodiert. Der Grund dafür ist, dass der Homesite-Editor nur die korrekte Eingabe von ANSI-Zeichen unterstützt. Das zeigt sich schon während der Eingabe: schaltet man in die Browser-Ansicht, zeigt sich der Fehler. Folglich macht es auch keinen Sinn, eine so erstellte Datei abzuspeichern, die Zeichen einer so editierten und abgespeicherten Datei werden weder im Browser noch im Notepad und auch nicht im Homesite-Editor selbst richtig dargestellt.

In so fern muss man leider ohne eine brauchbare Unicode-Eingabe in Homesite leben.

Der "Trick" von Macromedia

Es gibt allerdings einen Ausweg bzw. mehr einen Umweg, den [Macromedia selbst beschreibt](#).

Zunächst aktiviert man die Unicode-Speichern-Unterstützung unter Options/Settings/File Settings "*Enable non-ASCII file encoding*". Die HTML-Editierung macht man in Homesite. Dann speichert man die Datei mit der Standard-Unicode-Kodierung in Homesite und öffnet sie daraufhin parallel in Notepad. Denn

die Unicode-Eingabe kann nur in Notepad korrekt gemacht werden (deutsche Umlaute scheinen auch in Homesite in Unicode zu funktionieren). Das Dokument speichert man dann anschließend in Notepad und lädt es in Homesite neu, wobei dieser letzte Schritt eigentlich überflüssig ist. Denn in der Browser-Ansicht von Homesite werden die Zeichen zwar korrekt angezeigt, doch der Homesite-Editor selbst stellt die Zeichen weiterhin fehlerhaft dar.

Hinzu kommt, dass man kein einziges weiteres Zeichen über den Homesite-Editor eingeben darf, das führt nämlich wieder zur besagten falschen Kodierung, selbst dann, wenn man ein Zeichen eingibt und es gleich wieder löscht. Genauso wenig darf man die Datei in Homesite mit irgendeiner Kodierung, eben auch nicht mit Unicode, abspeichern, denn auch das korrumpiert die Kodierung.

Die Eingabe und Speicherung von deutschen Buchstaben in Homesite mit Unicode scheint zu funktionieren.

Der Trick mit Unicode UTF-8

Es gibt allerdings eine weitere und, wie ich meine, gar nicht so schlechte Lösung, die zumindest Abhilfe beim Editieren des Quelltextes und dem Abspeichern der Datei in Homesite schafft:

Dazu muss man ähnlich vorgehen, wie oben von Macromedia beschrieben, aber mit einem kleinen Unterschied: die Datei muss im UTF-8 Format abgespeichert werden. Eine UTF-8 kodierte Datei lässt sich in Homesite öffnen und auch korrekt speichern, vorausgesetzt man editiert die Datei in Homesite nur mit den im ANSI-Code festgelegten Zeichen, also die auf einer englischen Tastatur: a bis z, 0-9 etc., übliche Sonderzeichen, wie z.B. !\$%, aber keine Umlaute (übrigens auch kein §) und dergleichen. Natürlich dürfen auch nur solche Zeichen per "Kopieren und Einfügen" eingefügt werden.

Im Homesite-Editor werden die Zeichen zwar fehlerhaft dargestellt, kodiert werden sie aber richtig, was man in der Browser-Ansicht jederzeit kontrollieren kann. So kann man zumindest die HTML-Tags und was sonst noch zum Quelltext gehört editieren, ohne Schaden an der UTF-8 Kodierung der Datei anzurichten. Gibt man versehentlich Nicht-ANSI-Zeichen ein, kann man sie problemlos wieder löschen, was ja bei der Macromedia-Methode gar nicht möglich ist.

Achtung: Erst nach dem man die Datei nach dem Editieren speichert, werden alle Zeichen richtig im Browser angezeigt. Selbst wenn man Nicht-ANSI-Zeichen eingegeben und die Datei in Homesite versehentlich gespeichert hat, kann man die nicht erlaubten Zeichen noch löschen und die Datei retten, wenn sie im Homesite-Editor nach dem Speichern geöffnet bleibt. Nachdem man die Nicht-ANSI-Zeichen gelöscht hat, speichert man die Datei (hoffentlich) fehlerfrei und kann sie schließen.

Nachtrag: Ich habe festgestellt, dass es sogar möglich ist, Nicht-ANSI-Zeichen, beispielsweise deutsche Umlaute, einzugeben und die Datei darauf in Homesite im UTF-8-Format abzuspeichern. Die Datei kann daraufhin korrekt im Browser angezeigt und auch problemlos in Homesite geladen werden. Wichtig ist in diesem Zusammenhang, dass

1. es sich um die erste Eingabe von Nicht-ANSI-Zeichen handelt
2. nach der erstmaligen Eingabe und Abspeichern der Nicht-ANSI-Zeichen und dem darauf folgenden **Laden** der Datei keine weiteren Nicht-ANSI-Zeichen eingegeben und in der Datei gespeichert werden. Denn darauf werden die zuvor eingegebenen Nicht-ANSI-Zeichen korrumpiert.

Entscheidend ist also, dass die Datei nicht in Homesite nachträglich geladen, mit Nicht-ANSI-Zeichen gefüllt und abgespeichert wird. Denn solange nicht gespeichert wird, können Nicht-ANSI-Zeichen, die im Nachhinein (z.B. versehentlich) eingegeben wurden, einfach gelöscht werden. Jede Speicherung von neu eingegebenen Nicht-ANSI-Zeichen nach dem Laden einer UTF-8-Datei mit Nicht-ANSI-Zeichen zerstört die davor eingegebenen und gespeicherten Nicht-ANSI-Zeichen.

Gibt man beispielsweise folgende Buchstaben ein und speichert die Datei als UTF-8, so handelt es sich um eine intakte UTF-8-Datei:

Ü Ä Ü

Ladet man nun diese Datei wieder in Homesite und gibt noch zusätzlich ein Nicht-ANSI-Zeichen wie Ä ein, so gibt es bereits Probleme mit der Darstellung der Zeichen:

Ü Ä Ü Ö

Speichert man diese Datei ab, so korrumpiert man sie, die zuvor eingegebenen drei Buchstaben werden falsch kodiert, der neu eingegebene Buchstabe richtig.

Hinweis: Alle Angaben in diesem Artikel wurden von mir mit größter Sorgfalt erarbeitet. Die Arbeit erhebt aber keinen Anspruch auf Richtigkeit oder Vollständigkeit. Für eventuelle Schäden, die im Zusammenhang mit der Verwendung dieses Artikels stehen, übernehme ich keine Haftung.

Copyright © Ali Reza Razavi – Alle Rechte vorbehalten.
Dieser Artikel steht der Allgemeinheit zur Verfügung und darf in unverändertem Zustand weitergegeben werden. Eine Verwertung dieses Artikels, die über übliche Zitate hinausgeht, bedarf meiner ausdrücklichen Genehmigung.