TextPrep

Version 3.72 (0318)

ein ToolSet zur
Erzeugung von Wortlisten
und zur
Fehlerbereinigung von Quelltexten
für das
***Dragon* NaturalVocTool**

zur Erstellung von Fachvokabularen
für
***Dragon* NaturallySpeaking**
von
Nuance

Dr. Tilmann Bäßler
Juli 2009

Ein Produkt der Bein-EDV

Inhaltsverzeichnis

Einleitung

Im Benutzerhandbuch des Dragon NaturalVocTools wird eindringlich darauf hingewiesen, dass die Quelltexte zur Erstellung eines Fachvokabulars aufbereitet sein müssen, bevor sie als Input für das NaturalVocTool verwendet werden können.

Neben der Entfernung von Bildern, Tabellen und allen Formatierungsinformationen ist es vor allem wichtig,

- dass der Text fehlerfrei ist,
- dass man Wortlisten erstellt hat, z.B. :
    - Abkürzungen mit ihrer gesprochenen Form,
    - Fachausdrücke, die aus mehreren Worten bestehen,
    - mit Bindestrich verketteten Begriffe, bei denen der Bindestrich nicht gesprochen werden muss (z.B. „Mutter-Kind-Beziehung“),
    - Ausdrücke mit einem sog. Ergänzungsstrich (z.B. „Hals- und Beinbruch“, oder „Ankunftszeit und -ort“)
    - Maßeinheiten und sonstige Sonderworte mit ihrer gesprochenen Form.

Was leistet nun das Toolset **TextPrep**?

A) Quelltextbearbeitung

- Wandelt alle MS Word-kompatiblen Dateien (DOC, RTF,..) in TXT-Dateien um.
- Wandelt MS Excel-Dateien (XLS) in TXT-Dateien um.
- Wandelt PDF-Dateien in RTF-Dateien um.
- Zeigt alle Worte, Abkürzungen, usw. in ihrer Textumgebung, um ihre Bedeutung besser verstehen zu können.
- Zeigt den Quelltext in seiner gedruckten Form und parallel dazu die darin vorkommenden Sonderzeichen und nicht druckbaren Zeichen in ihrer Hexadezimalform. Dies erleichtert das Auffinden besonderer Zeichen bzw. Wortgebilde.
- Erlaubt, alle Dokumente mit Sonderzeichen und nicht druckbaren Zeichen von anderen Quelltexten zu trennen, um sie in einem nachfolgenden Schritt einheitlich bearbeiten zu können. So können z.B. alle Texte mit DOS ASCII-Zeichen bearbeitet werden, damit sie unter Windows ANSI die Umlaute korrekt darstellen.
- Es ist wichtig, Quelltexte in Gruppen von gleichem Textaufbau zu unterteilen, da dadurch störende Adress- und Namenselemente leichter entfernt werden können. **TextPrep** trennt auf einfache Weise die Quelldokumente nach ihrer Struktur anhand des Dokumentanfangs oder anhand von Ausdrücken innerhalb des Dokuments.
- Manche Textkorrekturen müssen beim Erstellen von Fachvokabularen immer wieder durchgeführt werden. **TextPrep** merkt sich diese Korrekturen und führt sie „auf Knopfdruck“ bei jedem neuen Quelltext erneut aus. So gelingt es z.B., einheitliche Abkürzungen anzubieten, oder Quelltexte aus der DOS-Zeit in Windows ANSI Texte umzusetzen.
- Durch Worttrennungen am Zeilenende entstehen im VocTool viele unerwünschte Wortfragmente. **TextPrep** fügt diese Wortteile wieder zusammen und „rettet“ diese und ihre Wortumgebung für das Vokabular.
- Benennt alle Quelldateien systematisch um, um evtl. Hinweise auf Patienten zu eliminieren.

- Fasst auf Wunsch alle Quelltextdateien zu einer einzigen oder mehreren großen Dateien zusammen.
- Ist dann schließlich der Quelltext mit TextPrep bearbeitet, ruft man das Nuance VocTool auf, um eine erste Liste aller unbekannten Worte zu erzeugen. Man stellt in aller Regel fest, dass doch noch sehr viele Worte mit Schreibfehlern vorhanden sind, die dann mühsam Wort für Wort mit Hilfe von z.B. MS Word und der Funktion „Ersetzen“ korrigiert werden müssen, um den Kontext für die Bi- und Trigrammstatistiken zu erhalten und ein fehlerfreies Vokabular nach einem zweiten Lauf von VocTool anbieten zu können. Auch hier erleichtert TextPrep das Umsetzen der Korrekturen: Mittels TextPrep bearbeitet man die Wortliste von VocTool mit einer Reihe von Hilfsmitteln, bis die Schreibweise der einzelnen Begriffe korrekt ist. Und damit ist die Arbeit auch schon getan: **TextPrep überträgt diese Änderungen selbsttätig auf den Quelltext!**
- Weitgehendes Umsetzen von Quelltexten alter deutscher Rechtschreibung in **neue deutsche Rechtschreibung!**
- Löscht unerwünschte Textabschnitte (Adressteile, Personennamen, Literaturstellen, usw.) am Anfang, am Ende, und beliebig auch mitten in den Quelltextdokumenten.

B) Erstellen von Wortlisten, z.T. mit der gesprochenen Form

- Erstellt eine Liste aller gesperrt geschriebenen Begriffe der Quelltexte. Diese können dann mit dem TextPrep-Tool REMOVERR in die Normalform überführt werden.
- Erstellt eine Liste aller Abkürzungen. In einem zweiten Schritt kann aus früheren Abkürzungslisten die gesprochene Form übertragen werden.
- Erstellt eine Liste aller Sonderbegriffe mit mehr als 1 Großbuchstaben (z.B. "ADAC", "MHz").
- Erstellt eine Liste aller Begriffe, die keinen Vokal enthalten. Dies sind sehr häufig Maßeinheiten (z.B. mg/ml, µg, °C).
- Erstellt eine Liste aller Mehrwortbegriffe (z.B. „Mutter-Kind-Beziehung“) **einschließlich ihrer gesprochenen Form** (d.h. ohne Bindestriche), damit die Bindestriche nicht mitdiktiert werden müssen. Neben dem Bindestrich lassen sich auch andere Trennzeichen wie z.B. „@“ auswählen, um so eine Liste aller Internetadressen zu erzeugen.
- Erstellt eine Liste aller Begriffe, die einen Ergänzungsstrich beinhalten z.B. „Hals- und Beinbruch“ oder „Knochenfrakturen und –behandlungen“ **einschließlich ihrer gesprochenen Form** (d.h. ohne Bindestriche).
- Erstellt eine Liste von aufeinanderfolgenden Begriffen, die groß geschrieben sind. Auf diese Weise werden z.B. Namen wie "Deutsches Rotes Kreuz" oder "Verband Deutscher Elektriker" gefunden.
- Adjektive oder Verben, die als Substantive gebraucht werden, werden von Dragon NaturallySpeaking oft nicht als solche erkannt und daher irrtümlich klein geschrieben. Beispiele: "beim Spielen", "etwas Schönes", "nach Abdecken und Abwaschen". TextPrep erzeugt eine Wortliste mit diesen Begriffen, um die korrekte Großschreibung sicherzustellen.

- TextPrep vervollständigt Fachausdrücke, deren erstes Wort in einer Liste angegeben ist.
  Beispiel: Es seien vorgegeben
  Arcus, Canalis, Corpus.
  Diese ergänzt TextPrep ganz oder teilweise zu
  Arcus plantaris profundus, Canalis semicularis, Corpus adiposum pararenale,
  sofern diese im Fließtext vorhanden sind.
  Die so ergänzten Ausdrücke können nun dem Vokabular hinzugefügt werden.

Installation

Systemvoraussetzungen

Da möglicherweise große Mengen von Text analysiert und bearbeitet werden müssen, gilt allgemein, je schneller der Rechner und je größer der Arbeitsspeicher, desto besser.

Mindestens sollte jedoch folgende Spezifikation erreicht bzw. übertroffen werden:

- 1,0 GHz Prozessor
- Windows 2000 oder Windows XP. Bei Windows 98 läuft das TextPrep-Tool SHOWHEX in seiner Darstellung nicht synchron (s. dort). Dieses Tool wird aber nur in besonderen Fällen bei „falschen“ Hexadezimalcodes von Sonderzeichen benötigt. TextPrep wurde unter Windows Vista nicht validiert.
- 512 MB RAM
- Etwa die vierfache Menge freien Plattenplatzes des zu bearbeitenden Quelltextes.
- Die Quelltextdokumente müssen im Windows ANSI-Format (TXT) vorliegen. Liegt das Quellmaterial in einem zu MS Word kompatiblen Format (.doc, .rtf etc.), oder im Format XLS (MS Excel)kann es mit dem Tool DOC2TXT von TextPrep zu .txt konvertiert werden. PDF-Dateien können zu RTF-Dateien umgewandelt werden.
- 1 USB-Port
- Ein zweiter Bildschirm als erweiterte Arbeitsoberfläche (von MS Windows XP unterstützt) ist zwar nicht unbedingt erforderlich, erleichtert das Bearbeiten von Wortlisten aber ganz erheblich; (s. Detailbeschreibung des Tools CONTEXT).
- Monitor, Mindestauflösung 1024 x 768.

Lieferumfang

- 1 TextPrep Programm-CD
- 1 USB-Dongle
- 1 CD Roche Rechtschreibprüfung Medizin

Installation von TextPrep

1. Windows starten.
2. Einlegen der TextPrep-Toolset CD in das CD-Laufwerk.
3. Öffnen des CD-Laufwerkordners mittels „Windows Explorer“ oder „Arbeitsplatz“.
4. Doppelklick im Unterverzeichnis "Programme" auf "TextPrep.msi".
5. Folgen Sie den Installationsanweisungen.
6. Kopieren der Datei "TextPrep-UserManual.pdf" von der CD in das TextPrep-Verzeichnis.

Hinweis: Sollte eine frühere Version von TextPrep installiert sein, so muss zuerst die alte Version deinstalliert werden. Die bisherigen TextPrep-Einstellungen gehen dabei nicht verloren.

Installation von Roche Rechtschreibprüfung Medizin

Das TextPrep-Tool **REMOVERR** verwendet das Fremdprogramm „Roche Rechtschreibprüfung Medizin“, um zusammen mit einer von Bein EDV mitgelieferten Wortlistendatei die korrekte Schreibweise von medizinischen Begriffen und Arzneimitteln anzuzeigen und ggf. zu übernehmen.

Die Original-CD der Roche Rechtschreibprüfung wird den Lizenznehmern von TextPrep kostenfrei zur Verfügung gestellt.

1. Einlegen der Original-CD Roche Rechtschreibprüfung Medizin.
2. Doppelklick auf setup.exe.
3. Auswählen: „Roche Rechtschreibprüfung und Lexikon“.
4. Auswählen: „Standard“.
5. Auswählen: „Medizin. Fachwörter (incl. Englische) und Fremdwörter“.
6. Auswahl der MS Word-Version: „nicht verwenden“.
7. Auswahl der Rechtschreibung: „neu, konservativ“.
8. Eingabe des Laufwerks + Unterverzeichnis (beliebig).
9. Nach der Programminstallation: Kopieren der Datei „medizin.txt“ von der Bein-EDV CD in das Roche-Unterverzeichnis.
10. Im Roche-Unterverzeichnis das Programm "rspell.exe" aufrufen.
11. „Einstellungen“ aufrufen
    - Hautwörterbuch: „Stichwortsuche des Roche-Lexikons“ auswählen.
    - Benutzerwörterbuch: Datei „medizin.txt“ auswählen
    - Exportwörterbuch: Datei „medizin.txt“ auswählen
12. Datei „rspell.dll von CD der Bein-EDV (nicht Roche-CD!) in TextPrep-Verzeichnis kopieren.
13. Datei „rspell.dll“ von CD der Bein-EDV (nicht Roche-CD!) in C:\windows\system32 kopieren.

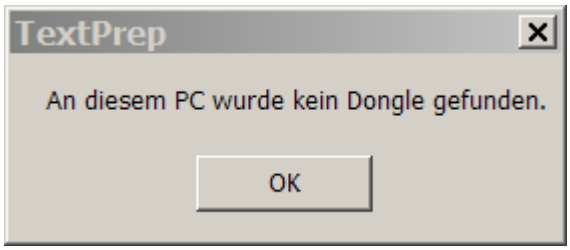
Wichtig: Die Pfadnamen der TextPrep-Unterverzeichnisse, besonders das Wortlistenverzeichnis, dürfen keine Sonderzeichen, insbesondere keine Umlaute ("ä", "ö", "ü") enthalten, da das Roche-Programm diese nicht erkennt. Die Rechtschreibprüfung zeigt in diesen Fällen keine alternativen Schreibweisen!

Aufruf / Beenden des ToolSets

Aufruf

Stecken Sie den USB-Dongle in einen USB-Port und Doppelklicken Sie auf das Programm-Icon , das man sich ggf. aus Bequemlichkeitsgründen auf die Windowsoberfläche oder in die Taskleiste kopiert hat.

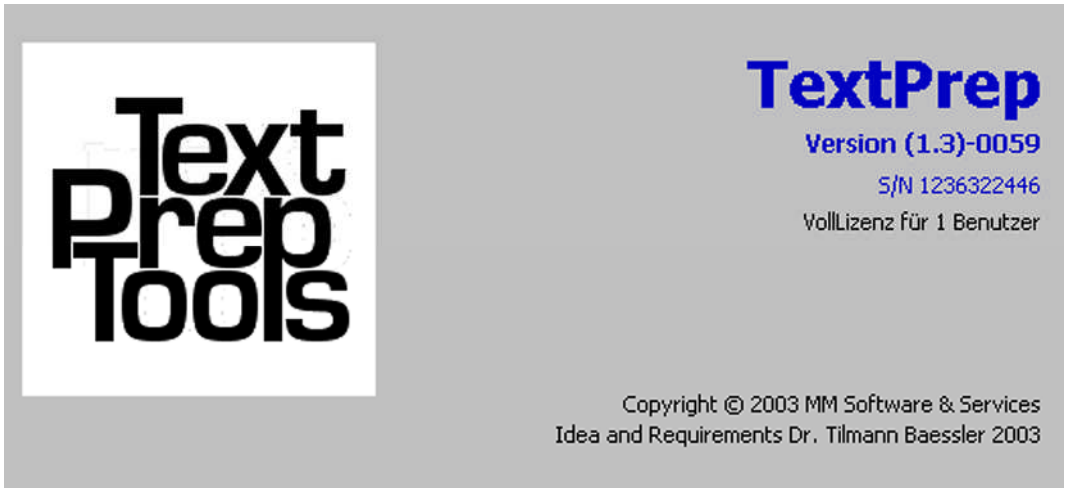
Erscheint beim Aufruf von TextPrep folgendes Fenster, ist dies ein Hinweis, dass kein Dongle gefunden wurde.



Stecken Sie den Dongle in einen freien USB-Port und starten Sie TextPrep erneut.

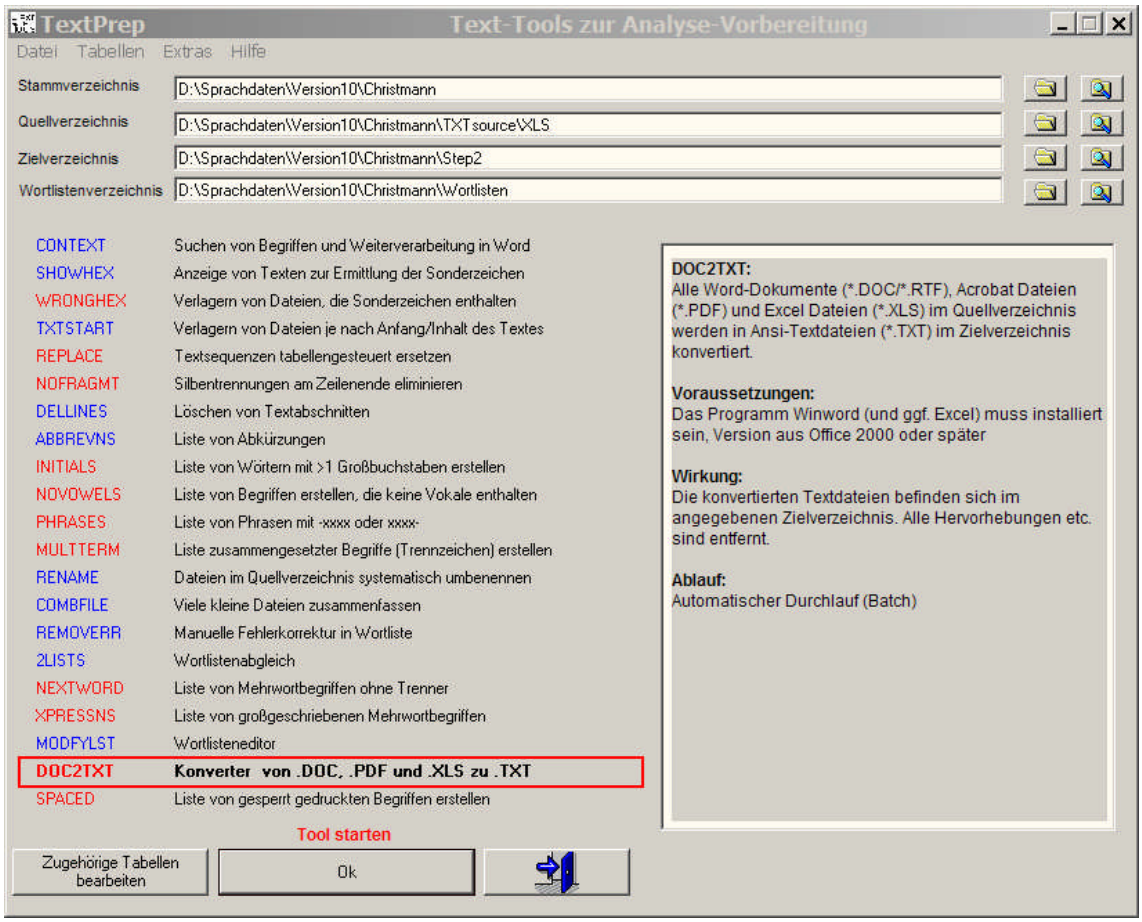
Beim allerersten Aufruf von TextPrep werden Sie aufgefordert, die Seriennummer einzugeben, die Sie zusammen mit der Programm-CD und dem USB-Dongle erhalten haben. Achten Sie bei der Eingabe der Seriennummer sorgfältig auf Groß- und Kleinschreibung! Der so aktivierte Dongle kann auch mit einem anderen Rechner verwendet werden, auf dem TextPrep installiert ist.

Ist die Seriennummer schon eingegeben, zeigt sich für einige Sekunden das Eingangsbild.



Diesem können Sie entnehmen, ob es sich um eine zeitlich unbegrenzte Vollversion handelt, oder im Falle einer Testversion, wie lange diese Kopie noch testweise genutzt werden kann. Sollten Sie sich zum Kauf einer endgültigen Lizenz entschließen, muss der Dongle nicht ausgetauscht werden. Durch Eingabe einer anderen Seriennummer, die Ihnen zugeschickt wird, lässt sich der Lizenzstatus des Programms ändern.

Auf das Eingangsbild erscheint nach einigen Sekunden das TextPrep-Hauptmenu:



Stammverzeichnis

Soll alle Unterverzeichnisse enthalten, die zur Erstellung eines Vokabulars angelegt werden. Dies erleichtert beim späteren Arbeiten die Auswahl des gewünschten Unterverzeichnisses.

Quellverzeichnis

Das ausgewählte Tool bearbeitet alle Dateien, die sich im Quellverzeichnis befinden und die vom lokalen MS Word geöffnet werden können (derzeit nur mit MS Word 2003 verifiziert).

Zielverzeichnis

Alle Dateien des Quellverzeichnisses werden nach der Bearbeitung durch das ausgewählte Tool im Zielverzeichnis gespeichert.

Wortlistenverzeichnis

Hier werden alle Wortlisten gespeichert, die durch TextPrep erstellt werden.

Die Tasten rechts neben den Verzeichnisfeldern erleichtern die Auswahl der jeweiligen Ordner.

Menuleiste

Datei

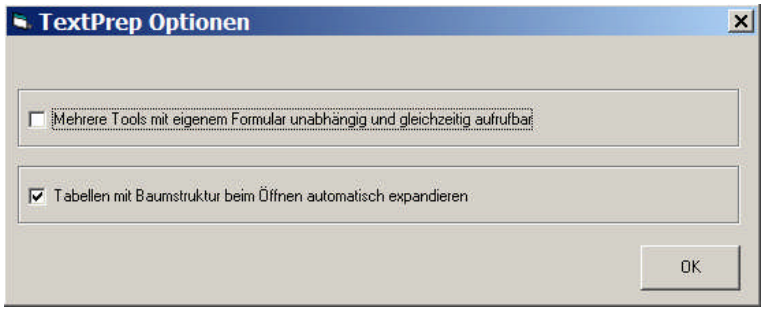
Bei diesem Menu lässt sich eine Datei „Einstellungen.txt“ in das Wortlistenverzeichnis exportieren bzw. von dort importieren. Diese Datei enthält alle Einstellungen und alle Tabellen, die bei den einzelnen Tools beschrieben werden. „Einstellungen.txt“ sollte möglichst nicht oder nur mit größter Vorsicht manuell geändert werden.

Tabellen

Die Tabellen ergänzen die jeweiligen Tools und erlauben eine hohe Flexibilität des Tooleinsatzes. Sie müssen zumindest vor dem erstmaligen Ausführen des Tools überprüft und ggf. angepasst werden. Sie werden ausführlich im Zusammenhang mit den einzelnen Tools beschrieben. Nach dem Aufruf eines Tools besteht die Möglichkeit, die zu diesem Tool zugehörigen Tabellen zu überprüfen bzw. zu bearbeiten.

Extras

Die hier enthaltenen **Optionen** erlauben,
a) TextPrep mehrmals parallel aufzurufen, und
b) die in TextPrep enthaltenen Tabellen in ihrer expandierten Form darzustellen



Hilfe

Unter diesem Menu wird bei „Lizenz“ der Lizenzcode eingegeben.

Ok

Mit einem Klick auf **Ok** starten Sie eines der angezeigten Tools, das Sie zuvor ausgewählt haben.

Beenden

Mit einem Klick auf das **Türsymbol**  beenden Sie das ToolSet TextPrep.

Farbgebung

Auf dem Hauptmenu sind die Namen der einzelnen Tools in **rot** oder **blau** wiedergegeben. Diese Farbgebung hat folgende Bedeutung:

rot Das Tool benötigt keine weiteren Angaben und beginnt nach dem Klicken auf die Taste „Ok“ sofort mit der Ausführung.

blau Es werden nach dem Klick auf die Taste „Ok“ noch weitere Angaben benötigt und erfragt, bevor das Tool ausgeführt werden kann.

Kurzbeschreibung der einzelnen Tools

CONTEXT

Alle Dateien im Quellverzeichnis werden nach einer Zeichenkette durchsucht. Die zu suchende Zeichenkette wird erfragt. Pro Datei wird wahlweise die erste Fundstelle mit Kontext angezeigt (pro Datei 1 Zeile) oder es werden alle Fundstellen angezeigt.
Dieses Tool ist hilfreich, um Abkürzungen und Sonderbegriffe in ihrer Bedeutung im Kontext zu verstehen bzw. bei der Bearbeitung von Wordlisten automatisch zu sehen, in welchem Kontext der markierte Ausdruck vorkommt.

SHOWHEX

Einzelne Texte im Quellverzeichnis können betrachtet werden. Dabei werden alle Sonderzeichen hexadezimal dargestellt.
Mit diesem Tool erkennt man, ob z.B. die Umlaute nicht im erforderlichen ANSI TXT-Format vorliegen, sondern im DOS ASCII-Format.

WRONGHEX

Dieses Tool arbeitet eng mit SHOWHEX zusammen. Alle in SHOWHEX gefundenen Dokumente, die „falsche“ Hex-Zeichen enthalten, können mit WRONGHEX extrahiert werden, damit in einem nachfolgenden Schritt (mit dem Tool REPLACE) diese Zeichen in die „richtigen“ ANSI-Zeichen umgesetzt werden können.

TXTSTART

Ein wichtiges Element der Dokumentaufbereitung ist die Anonymisierung der Quelltexte: Das zu erstellende Fachvokabular sollte weder Personennamen noch sonstige Adresselemente enthalten. Viele Quelltexte sind in ihrer Struktur einheitlich aufgebaut. Sie beginnen z.B. häufig mit dem Briefkopf.
Das Tool TXTSTART erlaubt, die Quelltexte nach ihrer Struktur zu trennen. Dateien mit einheitlicher Struktur können dann in einem nachfolgenden Schritt (mit dem Tool DELLINES) so bearbeitet werden, dass die Namens- und Adresselemente entfernt werden können.

REPLACE

Auf alle Dateien im Quellverzeichnis werden Ersetzungsvorschriften (d.h. Textkorrekturen) angewendet, die in der zugehörigen Tabelle festgelegt wurden. Alle Dateien des Quellverzeichnisses werden nach dem Ersetzungsvorgang im Zielverzeichnis gespeichert.
Dieses Tool kann sehr vielfältig eingesetzt werden. Es gibt gleich bleibende Ersetzungen, die beim Erstellen neuer Fachvokabulare jedes Mal vorgenommen werden müssen. Da bei TextPrep diese Ersetzungsanweisungen gespeichert bleiben, müssen sie nicht immer wieder neu eingegeben werden.
Beispiele:

- Die Umsetzung der DOS ASCII Umlaute in Windows ANSI Umlaute.
- Man sollte bestrebt sein, einheitliche Abkürzungen zu verwenden. Man findet z.B. als Abkürzung für das Wort „täglich“ sowohl „tägl.“ als auch „tgl.“. Man wird daher eine Ersetzungsanweisung eingeben, um alle Vorkommen von „tgl.“ in „tägl.“ umzusetzen.

NOFRAGMT

Ein Problem bei neuen Fachvokabularen ist leider, dass die gefundenen neuen Worte durch unerwünschte Wortfragmente verunreinigt sind, die durch die Silbentrennung am Zeilenende verursacht sind. NOFRAGMT erlaubt, diese Silbentrennungen weitgehend rückgängig zu machen. Es kann spezifiziert werden, in welchen Fällen dieses Rückgängigmachen nicht erfolgen soll.

DELLINES

Diese Tool ist wichtig für die Anonymisierung der Quelltexte. Es arbeitet eng mit dem Tool TXTSTART zusammen. DELLINES erlaubt wahlweise:

- die ersten x Zeilen aller Dateien im Quellverzeichnis zu löschen,
- den Anfang aller Dateien im Quellverzeichnis bis zu einer bestimmten Textstelle zu löschen,
- ab einer bestimmten Textstelle den Text bis zum Ende der jeweiligen Datei zu löschen,
- den Text zwischen zwei Textstellen zu löschen,
- in allen Dateien des Quellverzeichnisses diejenigen Zeilen zu löschen, in denen ein bestimmter Text (z.B. „Frau“ oder „Herr“) vorkommt,
- alle Zeilen zu löschen, in denen ein bestimmter Text nicht vorkommt. Der Nutzen dieser Funktion wird im Zusammenhang mit dem Tool REMOVERR erläutert.

ABBREVNS

Dieses Tool erzeugt eine Liste aller Abkürzungen, die in den Dateien im Quellverzeichnis vorkommen. Diese Abkürzungen werden in die Datei „Abkürzungen.txt“ im Stammverzeichnis eingetragen. In einem separaten Schritt ist dann die Liste der Abkürzungen manuell bzw. mit einem passenden TextPrep-Tool (2LISTS bzw. MODFYLST) halbautomatisiert mit der gesprochenen Form zu ergänzen.

INITIALS

Das Tool INITIALS erzeugt eine Wortliste aller Begriffe, die im Wort Großbuchstaben enthalten. Beispiele: ADAC, MHz, eMail. Diese Begriffe werden in die Datei „MehrGroßbuchstaben.txt“ im Stammverzeichnis eingetragen. In einem separaten Schritt ist dann die Liste der Begriffe manuell bzw. mit einem passenden TextPrep-Tool (2LISTS bzw. MODFYLST) halbautomatisiert mit der gesprochenen Form zu ergänzen.

NOVOWELS

Mit diesem Tool werden alle Begriffe gefunden, die keine Vokale enthalten. Damit sollen möglichst viele Maßeinheiten gefunden werden. Beispiele: mg/ml, µg, °C, qm. Diese Begriffe werden in die Datei „VokalfreieBegriffe.txt“ im Stammverzeichnis eingetragen. In einem separaten Schritt ist dann die Liste der Begriffe manuell bzw. mit einem passenden TextPrep-Tool (2LISTS bzw. MODFYLST) halbautomatisiert mit der gesprochenen Form zu ergänzen.

PHRASES

Um eine hohe Erkennungsrate zu erzielen, ist es für ein Fachvokabular sinnvoll, auch Begriffe wie z.B. „Hals- und Beinbruch“ oder „Knochenfraktur und -behandlung“ hinzuzufügen. Das Tool PHRASES extrahiert alle Ausdrücke dieser Art. Diese Begriffe werden in die Datei „Phrasen.txt“ im Stammverzeichnis eingetragen. Sie werden auf

Wunsch mit der gesprochenen Form automatisch ergänzt, damit diese Ausdrücke wie gewohnt ohne Ergänzungsstrich diktiert werden können.

MULTTERM

Sehr wichtig für Fachvokabulare sind Mehrwortbegriffe wie z.B. „Mutter-Kind-Beziehung“. Es soll dem Diktierenden erspart bleiben, die Bindestriche mitdiktieren zu müssen. Es ist daher sehr ratsam, diese Mehrwortbegriffe aufzufinden und in einer Wortliste mit der gesprochenen Form (ohne Bindestriche) bereitzustellen.
MULTTERM findet diese Begriffe, ergänzt sie auf Wunsch automatisch mit der gesprochenen Form, und speichert sie in der Datei „Mehrwortbegriffe“ in das Stammverzeichnis. Es kann festgelegt werden, ob auch Mehrwortbegriffe gefunden werden sollen, die ein anderes Sonderzeichen als der Bindestrich enthalten. So werden z.B. mit dem Zeichen „@“ alle E-Mail-Adressen gefunden.

RENAME

Manche Quelldateien tragen in ihrem Namen Hinweise auf Personen. Zur Anonymisierung ist es daher wünschenswert, diese Dateien umzubenennen. Das Tool RENAME ändert alle Dateinamen des Quellverzeichnisses in einen vorgegebenen Namen, ergänzt mit einer fortlaufenden Zahl.

COMBFILE

Dieses Tool fasst alle (TXT)-Dateien des Quellverzeichnisses zu 1 Datei oder einer frei wählbaren Anzahl von Dateien zusammen und legt sie im Zielverzeichnis unter dem Namen „GesamterText.txt“ ab.

REMOVERR

Die Liste neuer Wörter, die man z.B. mit Hilfe des Nuance VocTools oder einem der TextPrep-Tools erstellt hat, müssen in aller Regel überarbeitet werden Man korrigiert fehlerhafte Wörter, indem man mittels Auswahltasten die Schreibweise des Wortes ändert:

- nur der 1. Buchstabe ist groß geschrieben,
- alle Buchstaben sind groß geschrieben,
- das Wort wird klein geschrieben,
- die beiden Zeichen vor und nach dem Cursor werden getauscht,
- das Bindestrichwort wird zu 1 Wort zusammengefasst,
- das Bindestrichwort wird in 2 Worte getrennt,
- an der Stelle des Cursors wird das Wort in 2 Wörter mit Bindestrich getrennt mit wahlweise großgeschriebenem oder kleingeschriebenen 2. Wort
- Leerzeichen zwischen 2 Wörtern wird durch einen Bindestrich oder einem anderen, vorher spezifizierten Sonderzeichen ersetzt,
- das Wort(fragment) soll aus der Liste oder dem Quelltext entfernt (d.h. ersatzlos gelöscht) werden.
- Wahlweise wird der ausgewählte Begriff automatisch in die Windows Zwischenablage geschrieben. Dies erlaubt, wenn man das TextPrep-Tool CONTEXT parallel geöffnet hat, sofort die Textumgebung dieses Begriffes zu sehen.

Das markierte Wort befindet sich im Bearbeitungsfeld im Einfügemodus. Falls erforderlich, kann man es manuell beliebig verändern oder ergänzen.

Sind alle fehlerhaften Worte korrigiert, wählt man „Korrekturliste (auf Text) anwenden“, und Textprep überträgt automatisch alle Änderungen auf den ausgewählten Quelltext!

Das Tool REMOVERR wird ergänzt durch die Roche Rechtschreibprüfung Medizin. In einem separaten Fenster innerhalb des Tools werden zu dem aktuell ausgewählten Wort Alternativen gezeigt, die das Auffinden und Übernehmen der korrekten Schreibweise sehr erleichtern.

2LISTS

Bei den erzeugten Wortlisten **Bindestrichbegriffe** („Mutter-Kind-Beziehung“) und den **Begriffen mit einem sog. Ergänzungsstrich** („Auf- und Umbau“) wird die gesprochene Form auf Wunsch automatisch generiert.

Bei den Wortlisten **Abkürzungen** („Tabl.“)**, Maßeinheiten (**„mg/l“**)** und bei sog. **Großbuchstabenworte** („ADAC“) ist eine gesprochene Form unbedingt erforderlich, kann jedoch nur halbautomatisch erzeugt werden. Da diese Listen sehr lang sein können, ist es sehr zeitraubend, den einzelnen Wörtern die gesprochene Form manuell hinzuzufügen.

Es kommt noch hinzu, dass beim Erstellen von Fachvokabularen mehr als 80 Prozent dieser Begriffe und ihre gesprochene Form immer wieder vorkommen. Das Tool 2LISTS erlaubt nun, bei einer neu erstellten Liste die gesprochene Form aus früheren Listen auf leichte und sehr zeitsparende Weise zu übernehmen.

NEXTWORD

Adjektive und Verben werden oft im Zusammenhang mit den Wörtern "alles, etwas, nichts, …usw. als Substantive verwendet. Diese werden jedoch bei Dragon NaturallySpeaking oft klein geschrieben. Dies kann verhindert werden, wenn man diese Wörter als Begriff in das Vokabular aufnimmt, so z.B. "alles Schöne, nichts Gutes, etwas Erfreuliches, …usw.).

Umgekehrt möchte man auch Mehrwort-Ausdrücke wie z.B. "Diabetes mellitus" erfassen. NEXTWORD findet auch diese Begriffe ebenfalls.

XPRESSNS

Es gibt im Deutschen oft Namen von Organisationen usw., die aus mehreren großgeschriebenen Wörtern bestehen, ggf. auch mit kleingeschriebenen Füllwörtern. Beispiele: "Deutsches Rotes Kreuz", "Internationales Institut für Weltraumforschung". XPRESSNS erzeugt aus den Quelltexten eine Wortliste solcher Ausdrücke.

MODFYLST

Mittels dieses Tools kann man in einer Wortliste

- mittels MS Word und/oder einer anderen Rechtschreibprüfung die Wörter teilen in "richtig geschrieben" und "vermutlich falsch geschrieben". "Vermutlich falsch geschrieben" heißt nur, das dieses Wort der Rechtschreibprüfung unbekannt ist;
- die gesprochene Form erzeugen und/oder selektiv die gesprochene oder die geschriebene Form bearbeiten;
- auch bei z.B. englischen oder französischen Fremdwörtern die gesprochene Form erzeugen;
- ausgewählte Wörter und deren gesprochene Form in eine externe Datei auslagern;

- Textstrings am Anfang, am Ende oder irgendwo im Wort suchen und global ersetzen; sowohl in der geschriebenen oder in der gesprochenen Form;
- selektierte Wörter löschen;
- die gesprochene Form von Maßeinheiten automatisch generieren.
  Beispiel: mg/cm^2 → mg/cm^2\Milligramm pro Quadratzentimeter

DOC2TXT

Üblicherweise werden Quelldaten in einem MS Word-kompatiblen Format bereitgestellt. Da TextPrep nur mit dem eigentlichen Text ohne jegliche Formatierung arbeitet, muss man nahezu als allererstes die Quelldaten in das TXT-Format umsetzen. Dies lässt sicht leicht mit Hilfe des Tools DOC2TXT bewerkstelligen. Dieses Tool veranlasst MS Word, die Dateien aus dem Quellverzeichnis einzulesen und im TXT-Format in das Zielverzeichnis zu speichern. Die Größe der Dateien verkleinert sich auf weniger als 20 Prozent.

PDF-Dateien können ebenfalls in das TXT-Format umgesetzt werden.

Bei anderen Dateiformaten muss die Konversion in das TXT-Format ggf. manuell bzw. mit einer geeigneten Software vorgenommen werden.

SPACED

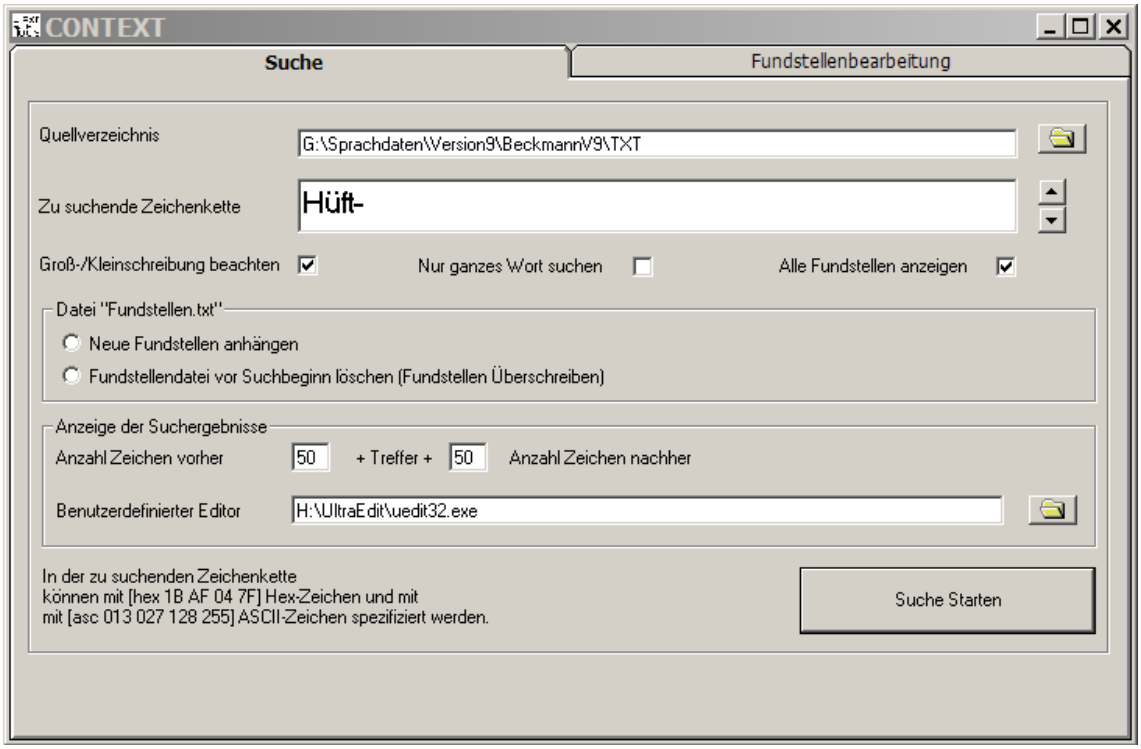
Dieses Tool listet in einer Datei alle Begriffe, die im Quelltext in gesperrt gedruckter Form vorkommen. Dies erlaubt mittels des TextPrep-Tools REMOVERR diese Begriffe in ihre normale Schreibweise zurückzuführen, um sie so den Wortanalysen bzw. -listen zugänglich zu machen.

Die Tools im Detail

A) Die Bearbeitung der Quelltexte

1. CONTEXT

Dieses Tool ist hilfreich, um Abkürzungen und Sonderbegriffe in ihrer Bedeutung im Kontext zu verstehen.
Zum Aufruf des Tools wählt man das Tool im Hauptmenu aus und bestätigt mit dem Auswahlknopf „OK“. Es erscheint das Fenster **Suche** des CONTEXT-Tools:



Im Feld **Quellverzeichnis** ist automatisch das Verzeichnis aus dem Hauptmenu übernommen worden.

Im Feld **Zu suchende Zeichenkette** wird der Suchbegriff eingegeben. Dieser Suchbegriff kann mit den Zeichen der Tastatur, in ASCII-Form, in hexadezimaler Darstellung oder aus einer Mischung von allem eingegeben werden. Dies ist dann erforderlich, wenn ein oder mehrere Zeichen der Zeichenkette nicht auf der Tastatur enthalten sind.
Um den Suchbegriff bei Bedarf leichter modifizieren zu können, kann der Suchbegriff und das Suchfeld mit den Dreieckstasten am rechten Rand vergrößert (bzw. verkleinert) werden.

Das Feld **Groß-/Kleinschreibung beachten** steuert die Suchgenauigkeit in Abhängigkeit von der eingegebenen Groß-/Kleinschreibung.

Die Option **Nur ganzes Wort suchen** vermeidet Treffer, bei denen die zu suchende Zeichenkette irgendwo innerhalb eines größeren Begriffs vorkommt.

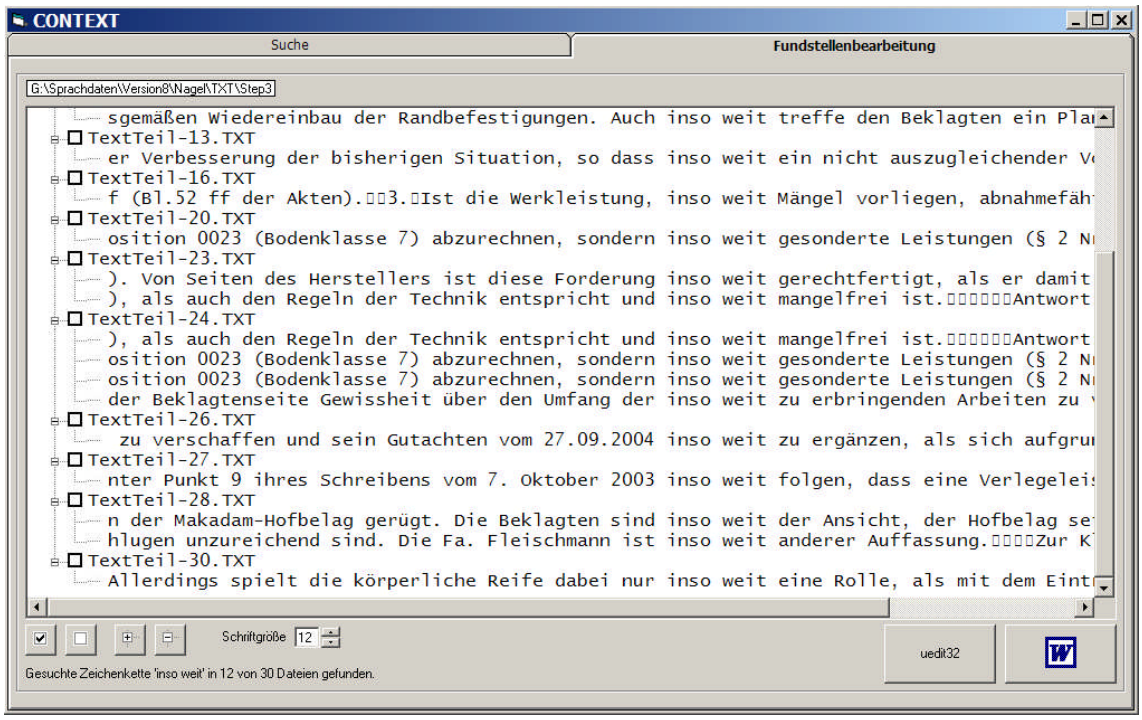
Die Option **Alle Fundstellen anzeigen** steuert, ob man pro Datei nur den ersten Treffer oder alle Treffer sehen möchte. Da bei der Option "alle Treffer" ggf. eine sehr lange Ergebnisliste entstehen kann, lässt sich der Suchvorgang anhalten. Man bekommt dann die bisher gefundenen Treffer angezeigt.

Das Suchergebnis, d.h. die Liste der Dateien, die den Suchbegriff enthalten, wird in der Datei „Fundstellen.txt“ im Wortlistenverzeichnis gespeichert, um die Dateien ggf. außerhalb von TextPrep bearbeiten zu können. Dabei kann man in den beiden Wahlfeldern angeben, ob das Suchergebnis an bisherige Suchergebnisse angehängt oder ob alte Suchergebnisse überschrieben werden sollen.

Bei der Anzeige der Suchergebnisse (eine Zeile pro Treffer) kann man angeben, wie viel Zeichen vor und wie viele nach dem Suchbegriff angezeigt werden sollen. Im abgebildeten Beispiel sind es jeweils 50.

Die Dateien des Suchergebnisses können auf Wunsch gleich mit Microsoft Word oder einem benutzerdefinierten Editor bearbeitet werden. Im Feld **Benutzerdefinierter Editor** kann man seinen bevorzugten Editor angeben.

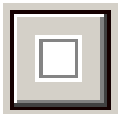
Mit einem Klick auf das Feld **Suche Starten** wird die Durchsuchung der Dateien des Quellverzeichnisses ausgelöst. Das Suchergebnis wird dann automatisch auf der Seite **Fundstellenbearbeitung** angezeigt:

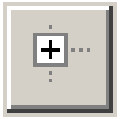


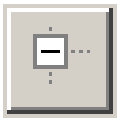
Man erkennt in der Mitte jeder Zeile den eingegebenen Suchbegriff „inso weit“ mit dem jeweiligen Kontext.

Zur Darstellung oder direkten Bearbeitung der Textstellen hat man folgende Optionen:


Alle Dateien markieren


Die Auswahlmarkierung bei allen Dateien löschen


Alle Treffer in den Dateien anzeigen


Nur den 1. Treffer pro Datei anzeigen.

Möchte man einzelne Dateien direkt mit dem Editor bearbeiten, markiert man sie in dem Feld links vom Dateinamen und wählt dann den gewünschten Editor.
Im abgebildeten Beispiel steht „uedit32“ stellvertretend für den Editor, den man auf der Seite zuvor (Reiter **Suche**) spezifiziert hat.

Zur besseren Lesbarkeit der Treffertexte kann die Schriftgröße beliebig eingestellt werden

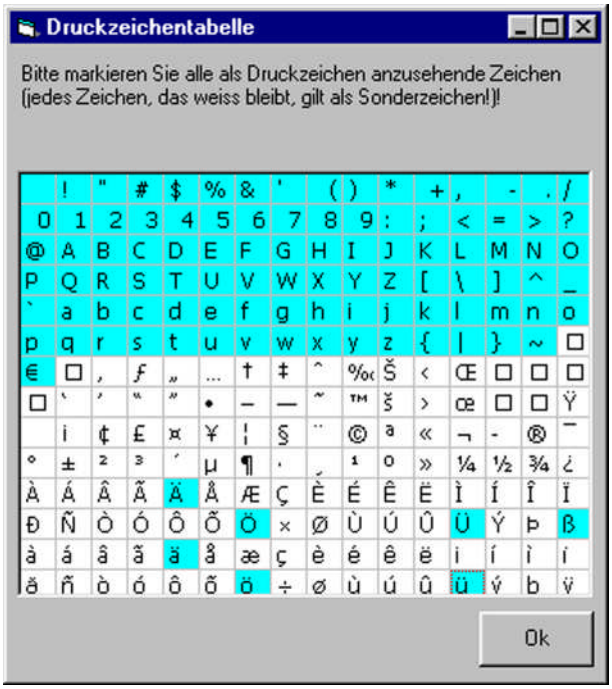
2. SHOWHEX

Mit diesem Tool kann festgestellt werden, in welcher hexadezimalen Codierung die Sonderzeichen im Text vorliegen. Welche Zeichen als Sonderzeichen und welche als „normale“ Zeichen anzusehen sind, wird durch die zugehörige Tabelle gesteuert.

a) Definition der Sonderzeichen

Nachdem man im Hauptmenu SHOWHEX markiert hat, klickt man (beim allerersten Aufruf dieses Tools) auf die Taste „Zugehörige Tabelle bearbeiten“.

Daraufhin wird eine Tabelle aller Zeichen angezeigt:

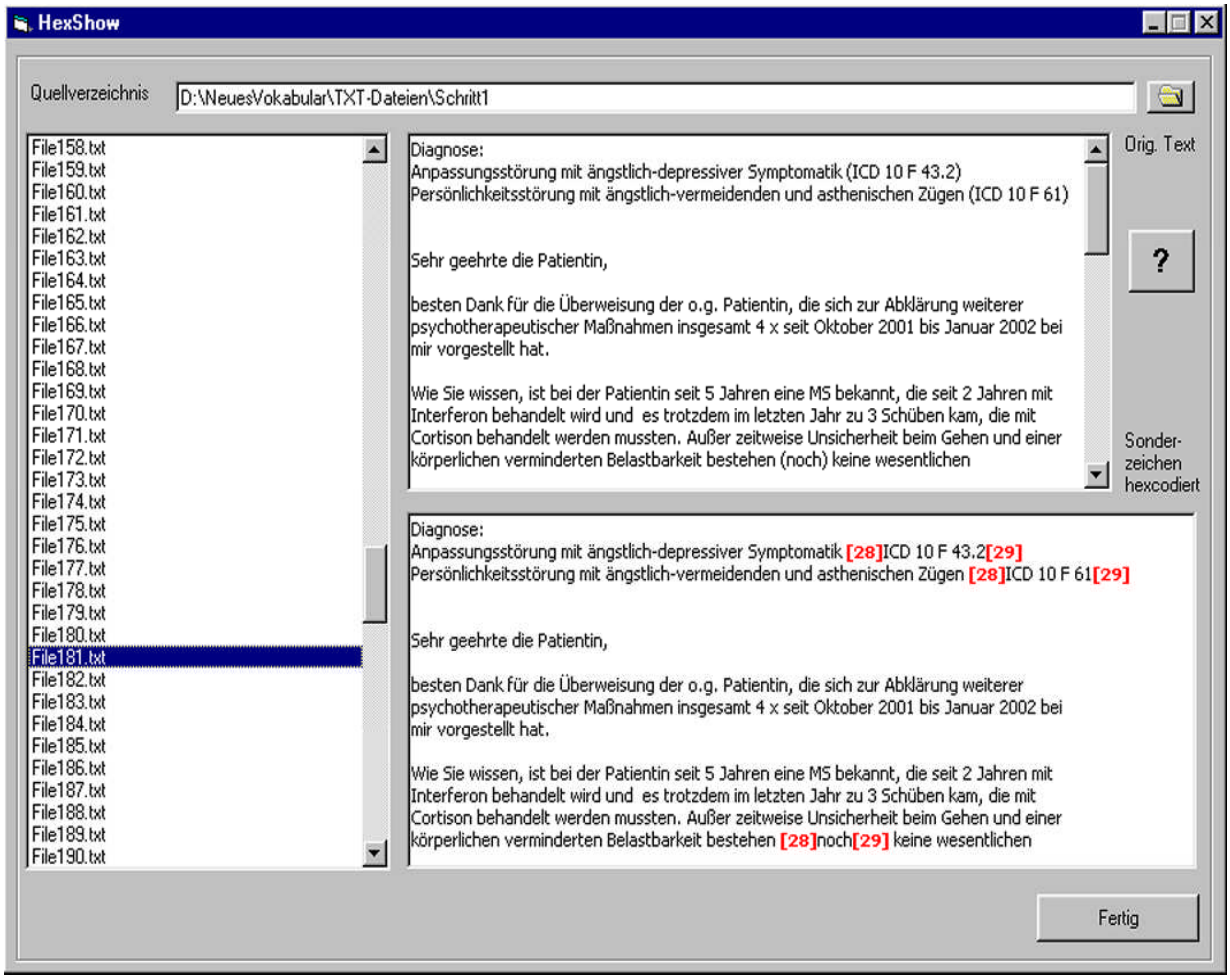


Alle schattierten Felder sind als „normale“ Zeichen ausgewählt, d.h. alle weißen Felder enthalten Sonderzeichen. Durch Anklicken eines Feldes ändert dieses seinen Status „normales Zeichen“ / „Sonderzeichen“.
Durch Anklicken von **Ok** schließt sich dieses Fenster wieder.

b) Darstellung der Sonderzeichen im Text in Hex-Code

Man wählt im Hauptmenu das Tool SHOWHEX aus und klickt auf **Ok**.

Es öffnet sich ein Fenster mit einer Liste der Dateien (linkes Fenster). Wählt man eine der Dateien aus, bekommt daraufhin den Textanfang dieses Dokument in den rechten beiden Fenstern zu sehen:



Im vorliegenden Beispiel sind die Zeichen „runde Klammer auf“ und „runde Klammer zu“ in Hex-Code dargestellt.

Mit diesem Tool kann man feststellen, ob „ungewöhnliche Zeichen“ im Text enthalten sind und ob diese bei den Wortlisten berücksichtigt werden müssen.

Wenn man im oberen rechten Fenster ein oder mehrere Zeichen mit dem Cursor markiert und dann auf das Fragezeichen 

klickt, wird dieses Zeichen bzw. werden alle markierten Zeichen in ihrer Hexform dargestellt.

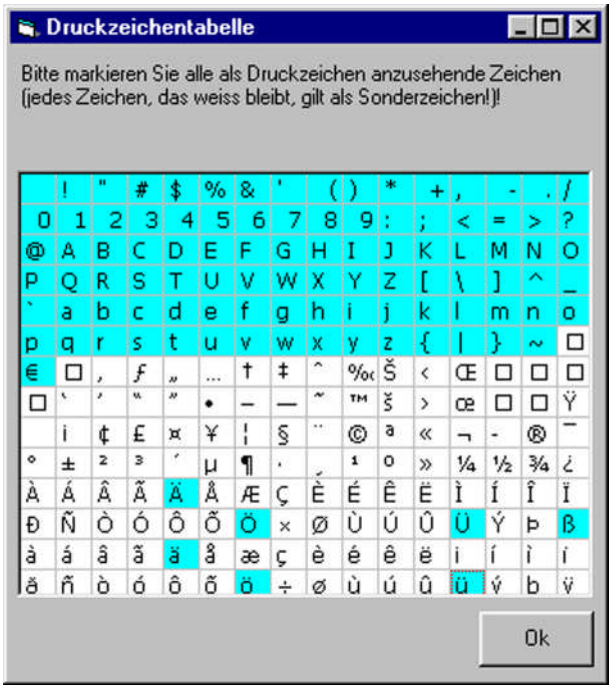
Mit einem Klick auf „Fertig“ schließt sich das Fenster des Tools SHOWHEX wieder.

3. WRONGHEX

Mit WRONGHEX können alle Dokumente extrahiert werden, die Sonderzeichen enthalten. Was Sonderzeichen sind, kann in der zu WRONGHEX zugehörigen Tabelle definiert werden.

a) Definition der Sonderzeichen

Nachdem man im Hauptmenu WRONGHEX markiert hat, klickt man (beim ersten Aufruf dieses Tools) auf die Taste „Zugehörige Tabelle bearbeiten“. Daraufhin wird eine Tabelle aller Zeichen angezeigt:



Alle schattierten Felder sind als „normale“ Zeichen ausgewählt, d.h. alle weißen Felder enthalten Sonderzeichen. Durch Anklicken eines Feldes ändert dieses seinen Status „normales Zeichen“ / „Sonderzeichen“.
Durch Anklicken von **Ok** schließt sich dieses Fenster wieder.

Ausführen des Tools WRONGHEX

Man wählt im Hauptmenu das Tool WRONGHEX aus. Dann überprüft man das Quellverzeichnis und das Zielverzeichnis und klickt dann auf **Ok**.

Das Tool verschiebt alle Dateien mit Sonderzeichen in das Zielverzeichnis. Dateien ohne Sonderzeichen bleiben im Quellverzeichnis!

Das Tool endet mit folgender, sinngemäßer Meldung:

Es zeigt, wie viel Dateien verlagert wurden. Mit Klick auf **OK** schließt man dieses Fenster.

In einem nachfolgenden Schritt können diese Dateien nun einheitlich bearbeitet werden. So kann man z.B. DOS ASCII Umlaute in die erforderlichen Windows ANSI Umlaute (mit dem Tool REPLACE) umsetzen.

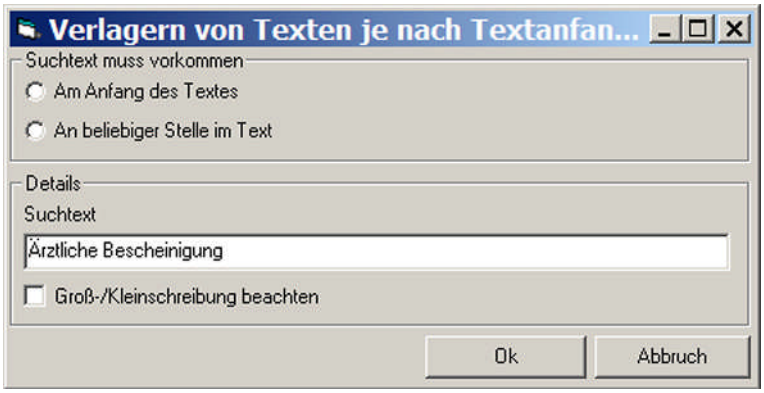
4. TXTSTART

Ein wichtiges Element der Dokumentaufbereitung ist die Anonymisierung der Quelltexte: Das zu erstellende Fachvokabular sollte weder Personennamen noch sonstige Adresselemente enthalten. Viele Quelltexte sind in ihrer Struktur einheitlich aufgebaut. Sie beginnen z.B. häufig mit dem Briefkopf.
Das Tool TXTSTART erlaubt, die Quelltexte nach ihrer Struktur zu trennen. Dateien mit einheitlicher Struktur können dann in einem nachfolgenden Schritt (mit dem Tool DELLINES) so bearbeitet werden, dass die Namens- und Adresselemente entfernt werden können.

Ausführen des Tools TXTSTART

Man wählt im Hauptmenu das Tool TXTSTART aus. Dann überprüft man das Quellverzeichnis und das Zielverzeichnis und klickt dann auf **Ok**.

Das Tool erfragt in einem separaten Fenster nach dem Textstring, den die Dokumente enthalten sollen:



Im Falle dieses Beispiels ist es: „Ärztliche Bescheinigung“.

Man hat nun die Möglichkeit, alle Dokumente auszuwählen, bei denen der Suchbegriff am Anfang des Dokuments vorkommt oder irgendwo im Text. Entsprechend klickt man eine der beiden Wahlmöglichkeiten an. Zusätzlich kann man vorgeben, ob die Groß-/Kleinschreibung beachtet werden soll.

Nach einem Klick auf **OK** verschiebt TXTSTART alle Dateien des Quellverzeichnisses mit dem Suchbegriff „Ärztliche Bescheinigung“ in das Zielverzeichnis. Alle anderen Dateien bleiben im Quellverzeichnis!

Das Tool meldet das Ende der Verarbeitung sinngemäß mit folgendem Text:

Mit Klick auf **OK** kann man dieses Fenster schließen.

In einem nachfolgenden Schritt können diese Dateien nun separat bearbeitet werden. So kann man z.B. Briefköpfe mit dem Tool DELLINES entfernen.

5. REPLACE

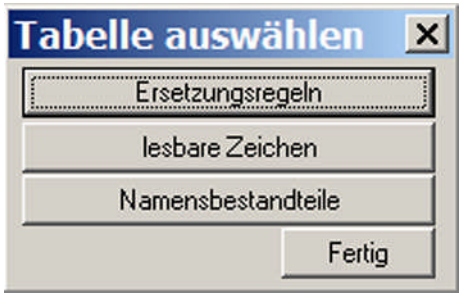
Dieses Tool ist besonders geeignet, Textkorrekturen und –umsetzungen vorzunehmen, die bei jeder Erstellung von Fachvokabularen immer wieder vorgenommen werden müssen.

Bevor das Tool aufgerufen wird, ist es erforderlich, Ersetzungsregeln einzugeben bzw. bereits vorhandene Ersetzungsanweisungen zu kennzeichnen, dass sie angewendet werden sollen.

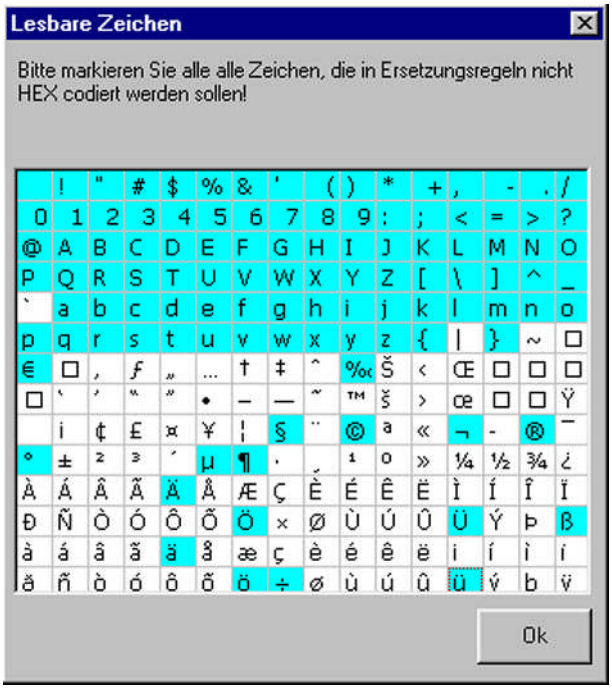
Wichtig: Mit REPLACE werden immer auch Teile eines Wortes ersetzt!! Ersetzt man z.B. "Herr" durch "Patient" entsteht aus dem Wort "Herrenberg" das Wort "Patientenberg". Ersetzungsanweisungen sind daher immer mit großer Sorgfalt zu erstellen.

a) Ersetzungsregeln bearbeiten

Nachdem man im Hauptmenu REPLACE markiert hat, klickt man (einmalig) auf die Taste „Zugehörige Tabellen bearbeiten“:

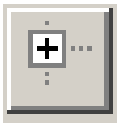


Wählt man **Lesbare Zeichen**, kann man in der folgenden Tabelle festlegen, welche Zeichen in den Ersetzungsregeln „lesbar“ dargestellt werden. Allen anderen Zeichen werden im ihrem Hexcode dargestellt.
Hinweis: Man sollte nur sehr wenige Zeichen als „nicht lesbar“ markieren, da sonst die Ersetzungsanweisungen ggf. sehr schwer lesbar werden.

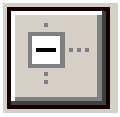


Wählt man **Ersetzungsregeln**, sieht man im folgenden Fenster die Ersetzungsregeln bzw. Gruppen von Ersetzungsregeln.

Aus Gründen einer besseren Übersicht sollten die Ersetzungsanweisungen (Regeln) in Gruppen eingeteilt werden. Ist in einzelnen Gruppen mind. 1 Ersetzungsanweisung aktiv, d.h. markiert, enthält das entsprechende gelbe Ordnerzeichen einen schwarzen Haken. (In diesem Beispiel der Ordner "Nagel").



Mit dieser Option werden alle Gruppen expandiert angezeigt, d.h. man sieht alle Ersetzungsanweisungen auf einen Blick.

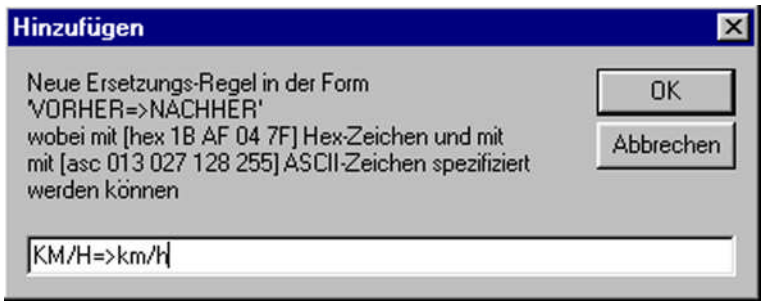


Mit dieser Option werden nur die Gruppen angezeigt, jedoch keine Ersetzungsanweisungen.

Damit lassen sich Ersetzungsanweisungen (nicht Gruppen!) nach oben oder unten verschieben.

Neue Gruppe: Es wird eine neue Gruppe angelegt. Man wird aufgefordert, einen Gruppennamen anzugeben.

Neue Regel: Es wird eine neue Regel eingegeben. Folgendes Eingabefenster erscheint:



Die Ersetzungsanweisung wird in folgender Form eingegeben:

alte_ Zeichenkette=>neue_Zeichenkette

In diesem Beispiel wird „KM/H“ durch „km/h“ ersetzt.
Die neue Ersetzungsanweisung erscheint in der Gruppe, die gerade aktiv ist. Die alte bzw. neue Zeichenkette kann auch (teilweise) hexadezimal eingegeben werden.

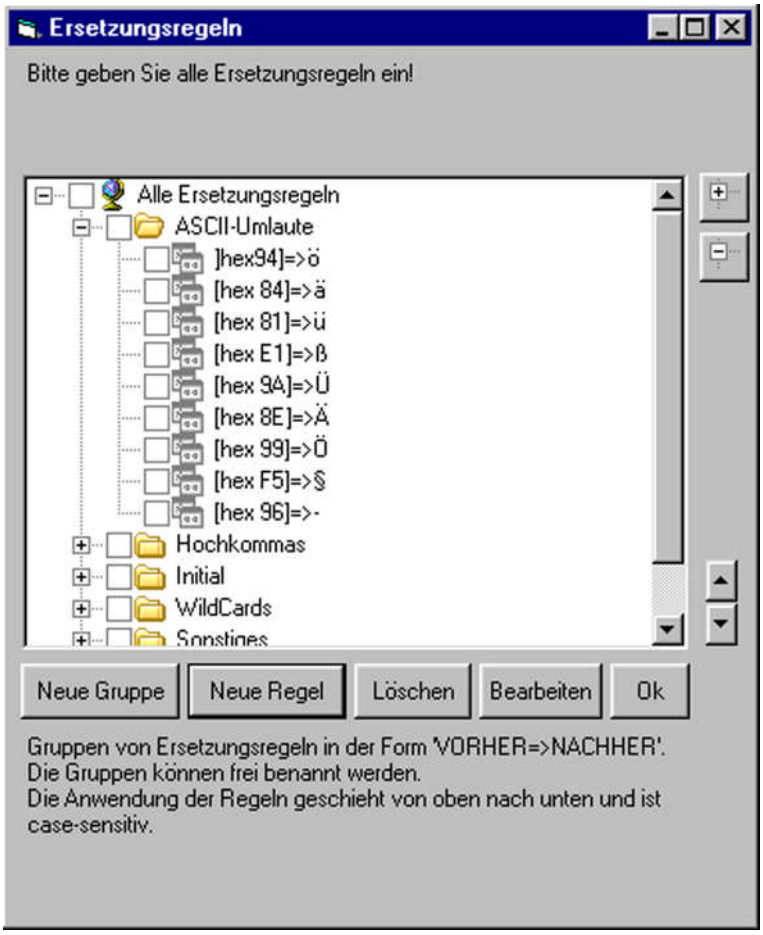
Löschen: Die markierte Gruppe oder Ersetzungsanweisung wird gelöscht.

Bearbeiten: Die markierte Gruppe kann umbenannt werden oder die markierte Ersetzungsanweisung kann editiert werden.

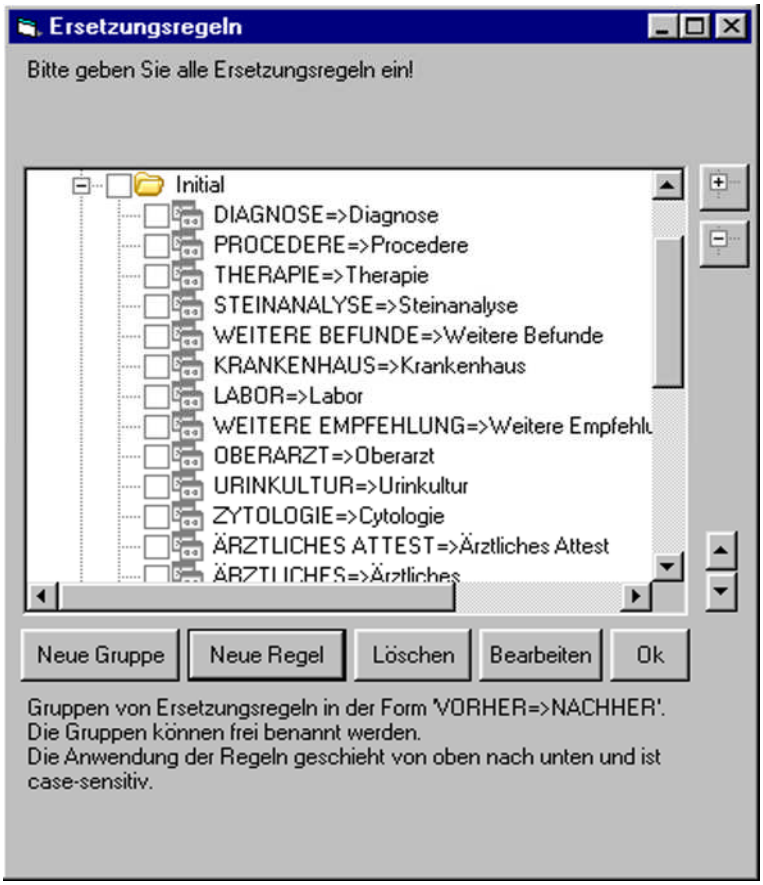
Ok: Die Ersetzungsregeln werden gespeichert.

Beispiele für Ersetzungsanweisungs-Gruppen:

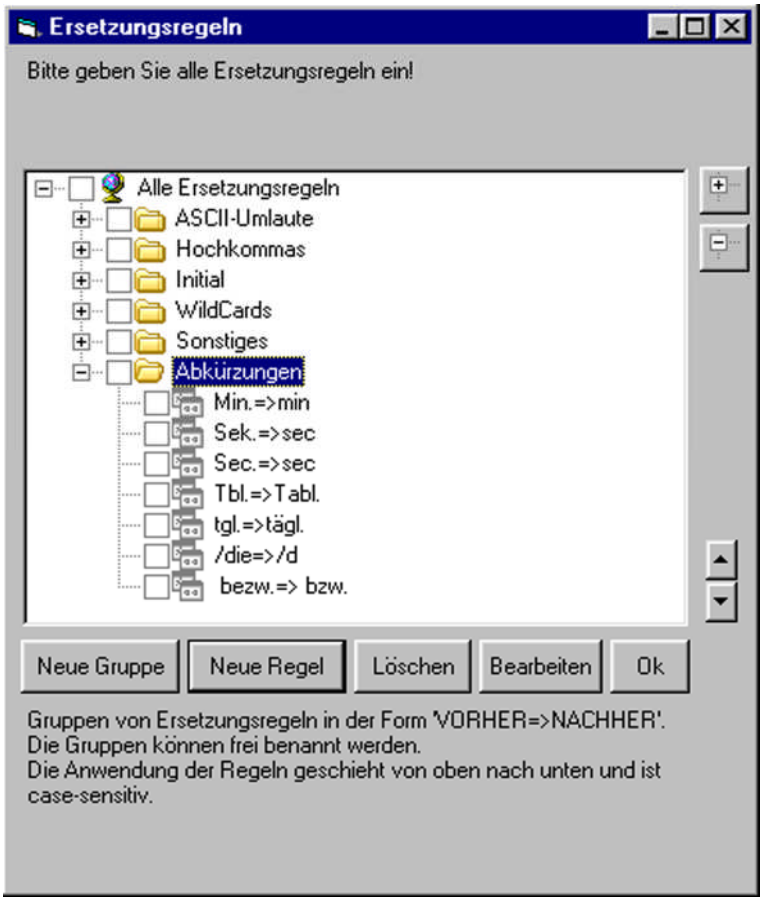
a) Umsetzung der Umlaute von DOS ASCII in Windows ANSI



b) Umsetzung von Wörtern, die in Großbuchstaben vorkommen und mit dem Tool INITIALS gefunden wurden.



c) Beispiel, wie man Einheiten und Abkürzungen auf eine normierte Schreibweise bringen kann:



d) **Die besondere Ersetzungsanweisung [w]**

Personennamen sind dann schwierig zu entfernen, wenn sie irgendwo im Fließtext, meist in der Form „Frau Mustermann“, „Herr Jedermann“, vorkommen. Mit der folgenden besonderen Ersetzungsregel kann man ein bestimmtes Wort und alle nachfolgenden großgeschriebenen **W**örter durch eine neue Zeichenkette ersetzen. Zusätzlich kann man in der zugehörigen Tabelle „Namensbestandteile“ Namenselemente (z.B. „Dr.“, „Ing.“, „Prof.“, „von“, usw.) eingeben, die ebenfalls gelöscht werden sollen. Damit soll erreicht werden, dass im Fließtext alle Namen von Personen anonymisiert werden.

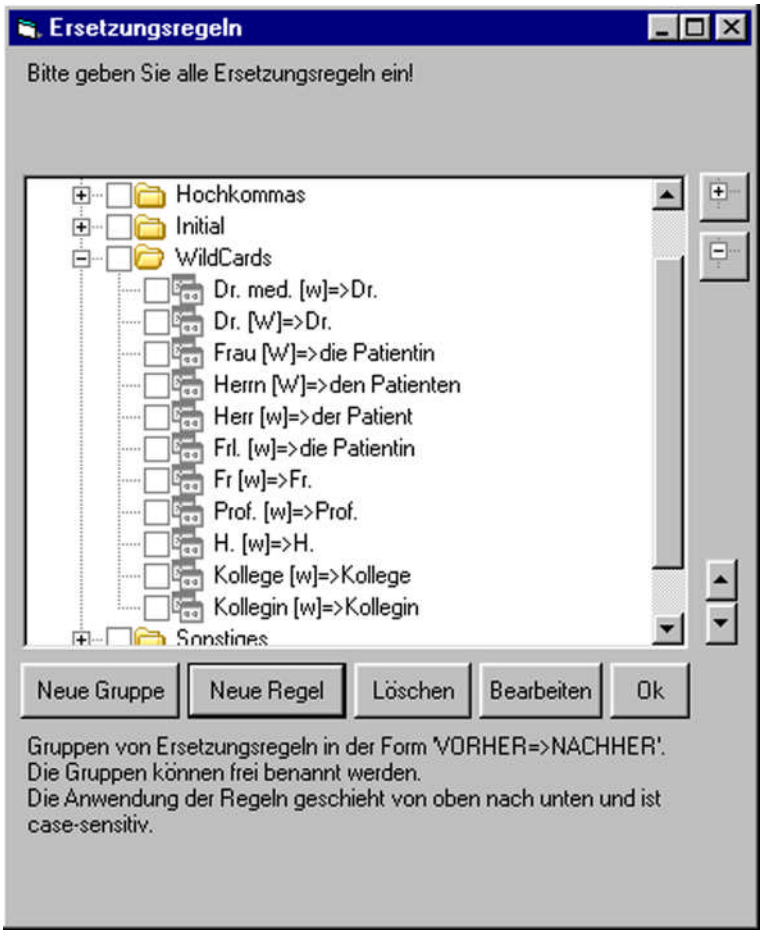
Beispiele: ….Frau Michaela Mustermann, ….Herr Prof. Dr. W. Jedermann, ….Graf Adelmann von der Teck, usw. Abgekürzte Vornamen (z.B. „S.“, „H.-J.“) werden ebenfalls als Namensbestandteil erkannt.

Die Ersetzungsanweisung wird in folgender Form eingegeben:

alte_ Zeichenkette[w]=>neue_Zeichenkette

In der folgenden Abbildung werden u.a. ersetzt:

- „Frau“ + die folgenden großgeschriebenen Worte durch „die Patientin“
- „Herrn“ + die folgenden großgeschriebenen Worte durch „den Patienten“, usw.



Achtung: Eventuelle Leerstelle(n) vor der Eckigen-Klammer-auf zählen mit zu der Zeichenkette "alte_Zeichenkette". Ggf. also keine Leerstelle vor der Klammer.

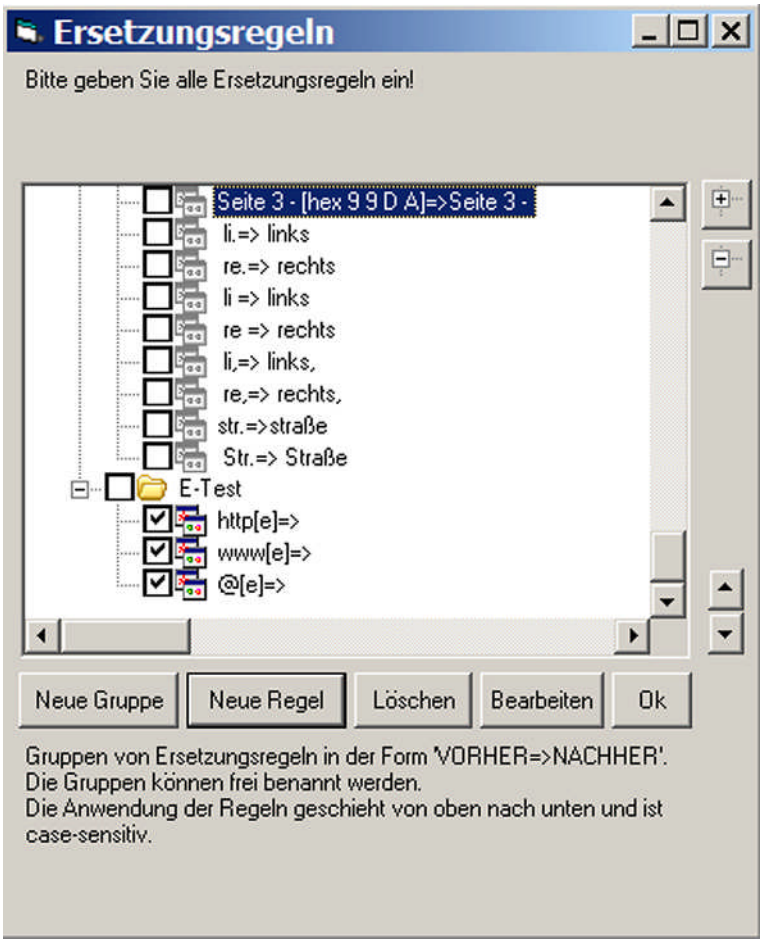
e) Die besondere Ersetzungsanweisung [e]

In Quelltexten treten vermehrt Internet- und Email-Adressen auf, die jedoch meistens nicht in einem Vokabular auftreten sollen. Um diese und ähnliche entfernen bzw. ersetzen zu können, setzt man die Ersetzungsregel „e“ ein.

Dieser Typ Ersetzungsanweisung erlaubt, alle diejenigen Wörter zu ersetzen, die eine bestimmte Zeichenkette **e**nthalten. Form der Eingabe:

alte_Zeichenkette[e]=>neue_Zeichenkette

In dem folgenden Beispiel werden alle Email-Adressen, gekennzeichnet durch das @-Zeichen, und alle Internet-Adressen, die die Elemente „http“ oder „www“ enthalten, durch eine Leerstelle ersetzt:



Besonderheit: Die Zeichenfolge **[e]** innerhalb der Ersetzungsanweisung bewirkt, dass die alte Zeichenkette nicht "case sensitive" ist, also die Groß-/Kleinschreibung innerhalb der Zeichenkette keine Rolle spielt.
Schreibt man hingegen **[E]**, muss die Groß-/Kleinschreibung exakt übereinstimmen, um zu einer Ersetzung zu führen.

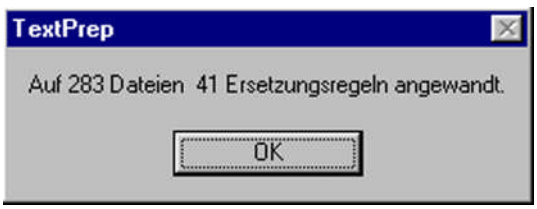
Hat man die zu REPLACE zugehörigen Tabellen **Ersetzungsregeln**, **Lesbare Zeichen** und **Namensbestandteile** bearbeitet, klickt man auf das Feld **Fertig**.

Ausführen des Tools REPLACE

Man wählt auf dem Hauptmenu das Tool REPLACE aus und klickt auf **Ok.**

Die mit einem Häkchen markierten Gruppen bzw. Ersetzungsanweisungen werden auf alle Dateien des Quellverzeichnisses angewendet. Alle Dateien werden in das Zielverzeichnis gespeichert.

Das Tool endet mit einer Meldung wie dieser:

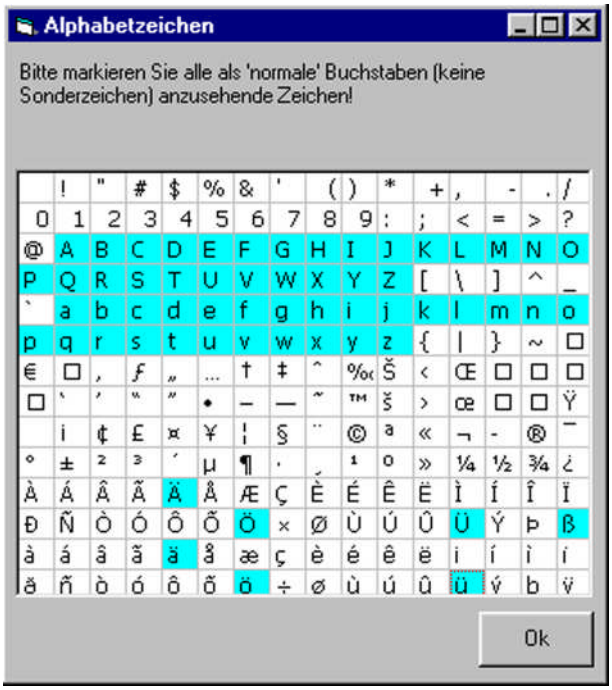


6. NOFRAGMT

Ein Problem bei neuen Fachvokabularen ist leider, dass die gefundenen neuen Worte durch unerwünschte Wortfragmente verunreinigt sind, die durch die Silbentrennung am Zeilenende verursacht sind. NOFRAGMT erlaubt, diese Silbentrennungen weitgehend rückgängig zu machen.

Zuerst muss definiert werden, was „normale“ Buchstaben und Zahlen sind.

Nachdem man im Hauptmenu NOFRAGMT markiert hat, klickt man (beim ersten Aufruf dieses Tools) auf die Taste „Zugehörige Tabelle auswählen“. Daraufhin wird eine Tabelle aller Zeichen angezeigt:



Man überprüft die ausgewählten (d.h. farbig markierten/schattierten) Zeichen und bestätigt mit einem Klick auf **Ok**. Die Tabelle wird daraufhin zurückgeschrieben.

Das Tool NOFRAGMT behandelt folgenden Fall:
Es sucht Wortgebilde, die am Zeilenende stehen und mit einem Trennstrich versehen sind. Zusätzlich muss zu Beginn der nächsten Zeile ein Wortgebilde aus „normalen“ Zeichen bestehen, von denen der erste kleingeschrieben sein muss.
Ist beides gegeben, fügt NOFRAGMT diese beiden Wortgebilde (ohne Trennstrich) zu 1 Wort zusammen.

Ausnahmen, bei denen die Wortgebilde nicht zusammengefügt werden:

- Am Anfang der 2. Zeile steht nicht ein „normales Zeichen, sondern ein oder mehrere Lehrstellen oder irgendein anderes (Sonder-)zeichen.
- Es gibt spezielle Ausnahmefälle. Beispiel:

  ----------irgendein Text ----------**Hals- und Beinbruch** ----------irgendein Folgetext----------

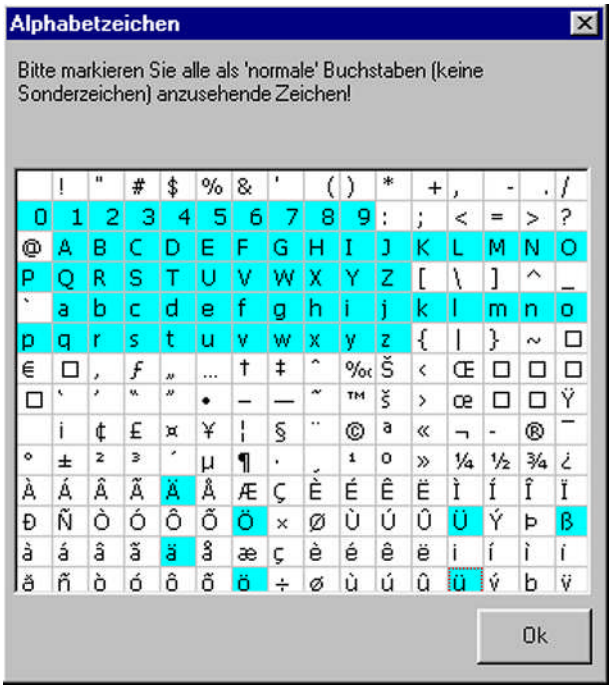
Obwohl „**Hals-**„ als auch „**und**“ den Kriterien für eine Zusammenfügung gehorcht, darf es natürlich nicht zu „Halsund“ zusammengefügt werden.

Es gibt daher eine Tabelle mit so genannten Stoppwörtern, bei denen eine Zusammenfügung nicht erfolgen soll.
Nachdem man im Hauptmenu NOFRAGMT markiert hat, klickt man auf die Taste „Zugehörige Tabellen auswählen“. Es werden zwei Tabellen angeboten.

Klickt man auf „**Sperrworte**“, wird eine Tabelle aller sog. Sperrworte angezeigt.

Man überprüft die Einträge, löscht ggf. einen Eintrag mit **Löschen** oder fügt einen zusätzlichen Eintrag mit **Neu** hinzu. Durch einen Klick auf **Ok** wird die Tabelle zurückgeschrieben.

Bei der Tabelle „**Alphabet**“ wird festgelegt, welche Zeichen in einem Wort vorkommen können.



Aufruf des Tools NOFRAGMT:
Im Hauptmenu von TextPrep markiert man das Tool NOFRAGMT, überprüft die Einträge im Quell- und Zielverzeichnis (ggf. korrigieren), und startet das Tool durch einen Klick auf **Ok**.

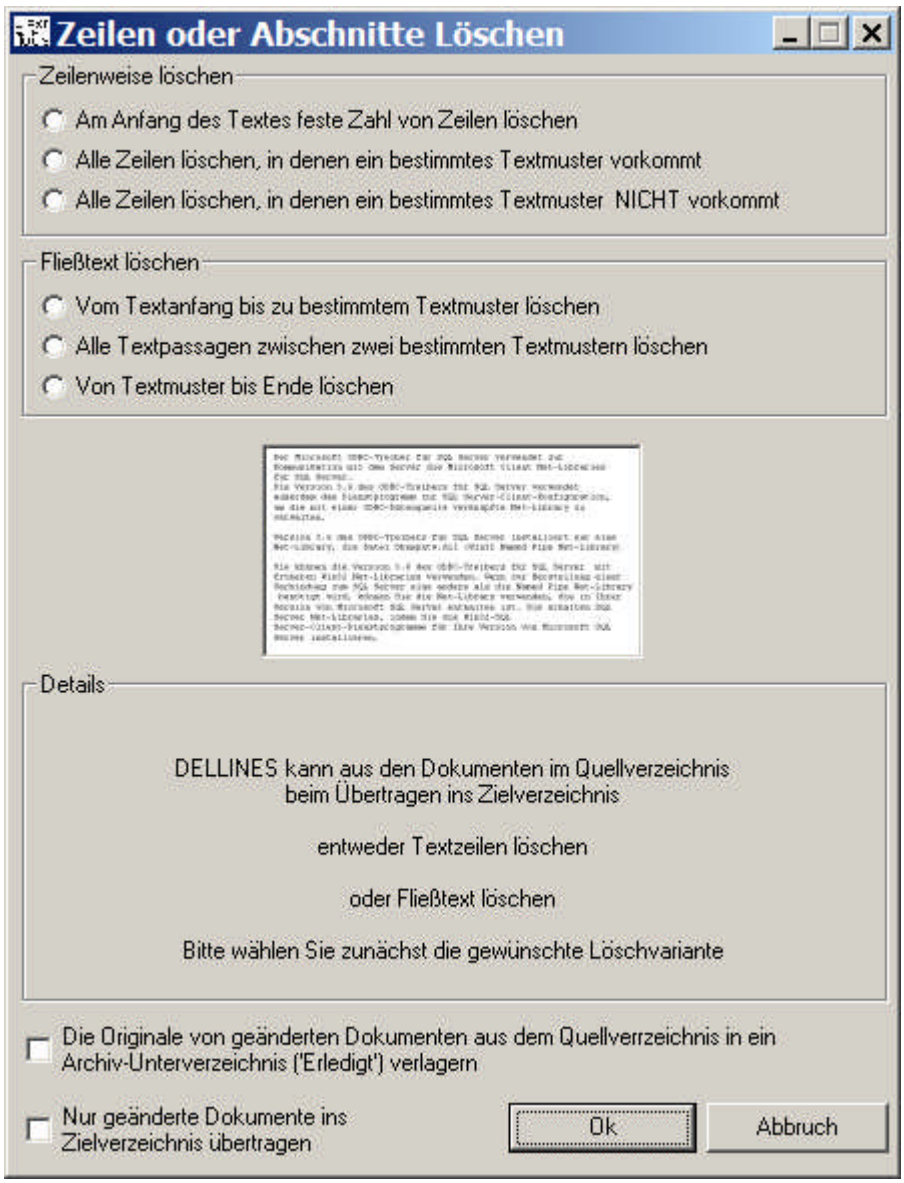
Das Tool behebt Zeilenumbrüche in allen Dateien des Quellverzeichnisses und schreibt alle Dateien in das Zielverzeichnis. Am Ende dieses Prozesses erscheint sinngemäß diese Meldung:

Mit Klick auf **OK** schließt sich dieses Fenster.

7. DELLINES

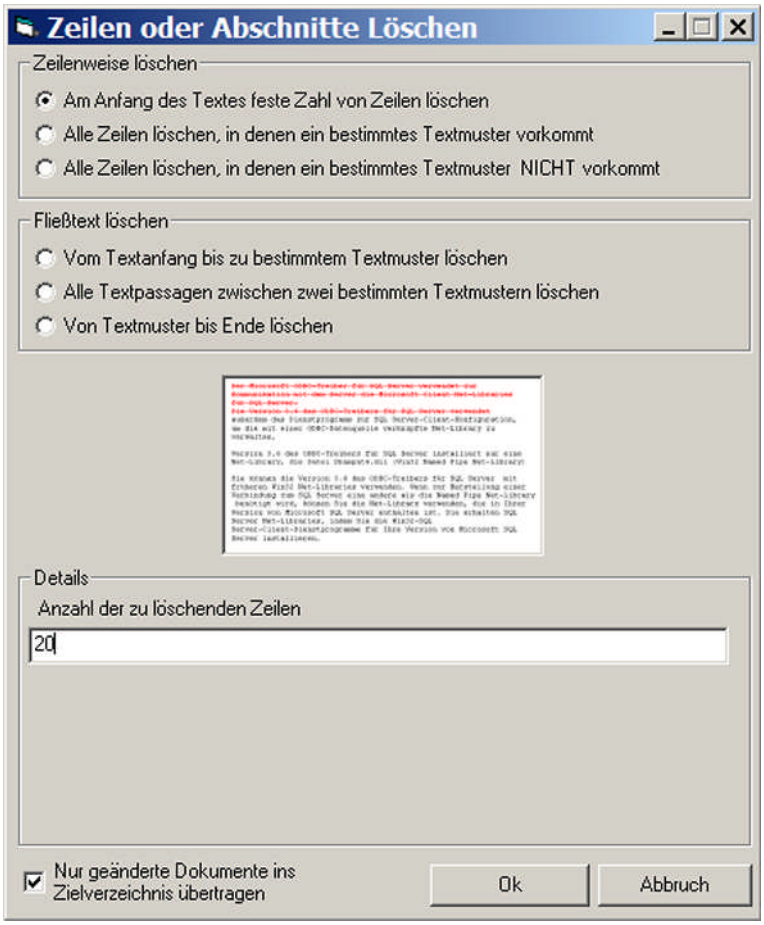
Dieses Tool ist wichtig für die Anonymisierung der Quelltexte. Es erlaubt, auf verschiedenste Weise Textpassagen (wie z.B. Adressblöcke) zu entfernen Es arbeitet eng mit dem Tool TXTSTART zusammen.

Durch Markieren des Tools DELLINES auf dem Hauptmenu von TextPrep und Klicken auf **Ok** öffnet sich folgendes Fenster:



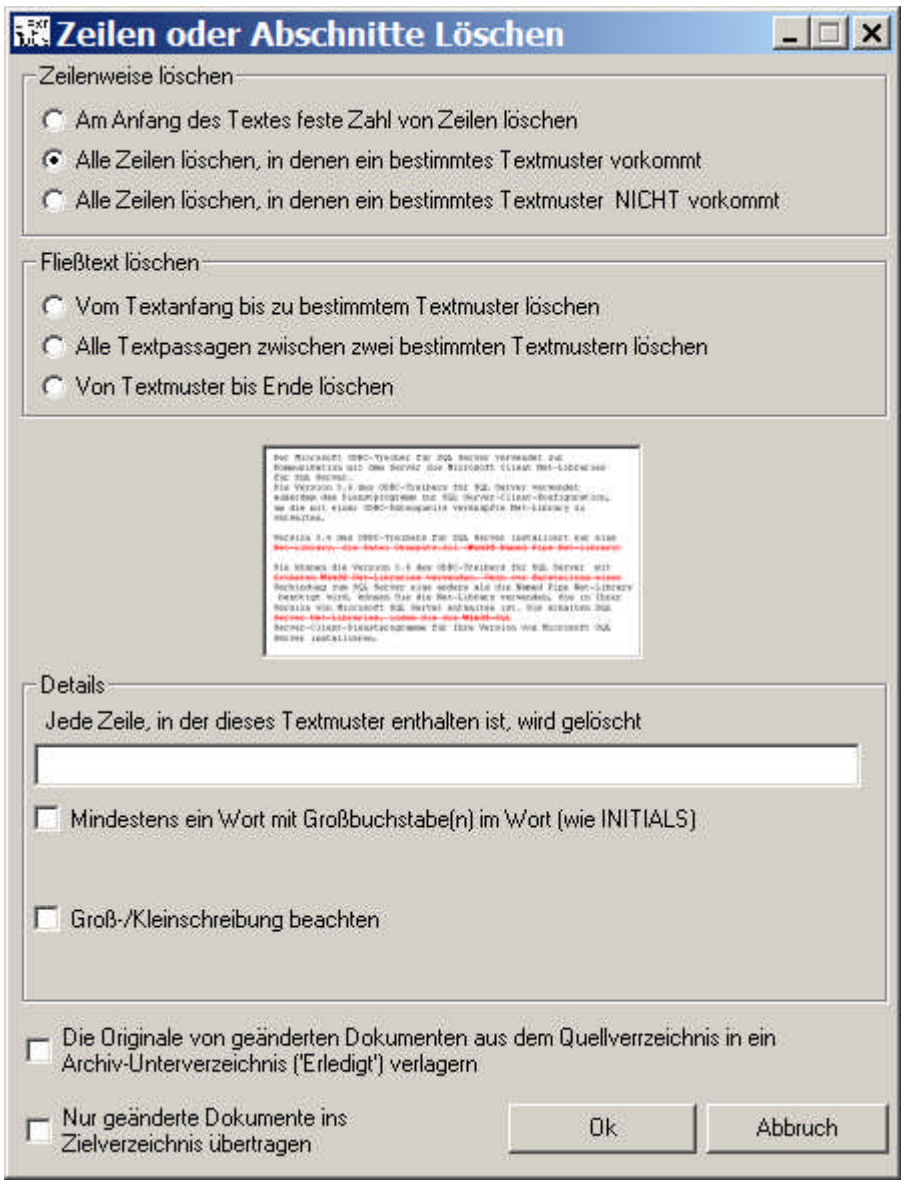
Es bieten sich mehrere Optionen:

a) Am Anfang jeder Datei des Quellverzeichnisses wird eine feste Anzahl von Zeilen gelöscht. Im folgenden Beispiel sind dies 20 Zeilen. Dies ist z.B. geeignet, um einen Briefkopf oder Adressteil fester Größe zu entfernen.



b) Jede Zeile im Text wird gelöscht, in der ein bestimmter Text vorkommt.

Im vorliegenden Beispiel werden alle Zeilen in allen Dateien des Quellverzeichnisses gelöscht, in denen die Zeichenkette „Patient:“ vorkommt.



Mit der Option „**Groß-/Kleinschreibung beachten**“ kann man die Suchgenauigkeit steuern.

Mit der Option "**Die Originale von geänderten Dokumenten aus dem Quellverzeichnis in ein Archiv-Unterverzeichnis ("Erledigt") verlagern**" kann man verfolgen, welche der Quelldokumente schon bearbeitet wurden und welche nicht.

Wenn im Fließtext Personennamen vorkommen, ist es oft schwierig, diese aufzuspüren und zu eliminieren. Da diesen Angaben oft „Frau“ bzw. „Herr“ vorangestellt ist, ist dies

eine Möglichkeit, diese Daten zu entfernen, indem man als Suchbegriff das Wort „Frau“ oder „Herr“ eingibt.

Eine besondere Option ist "**Mindestens ein Wort mit Großbuchstabe(n) im Wort (wie INITIALS)**. Ist sie ausgewählt, ist im Eingabefeld kein Textmuster erforderlich. Mit dieser Option kann man aus Wortlisten alle Initialwörter (ADAC …) entfernen.

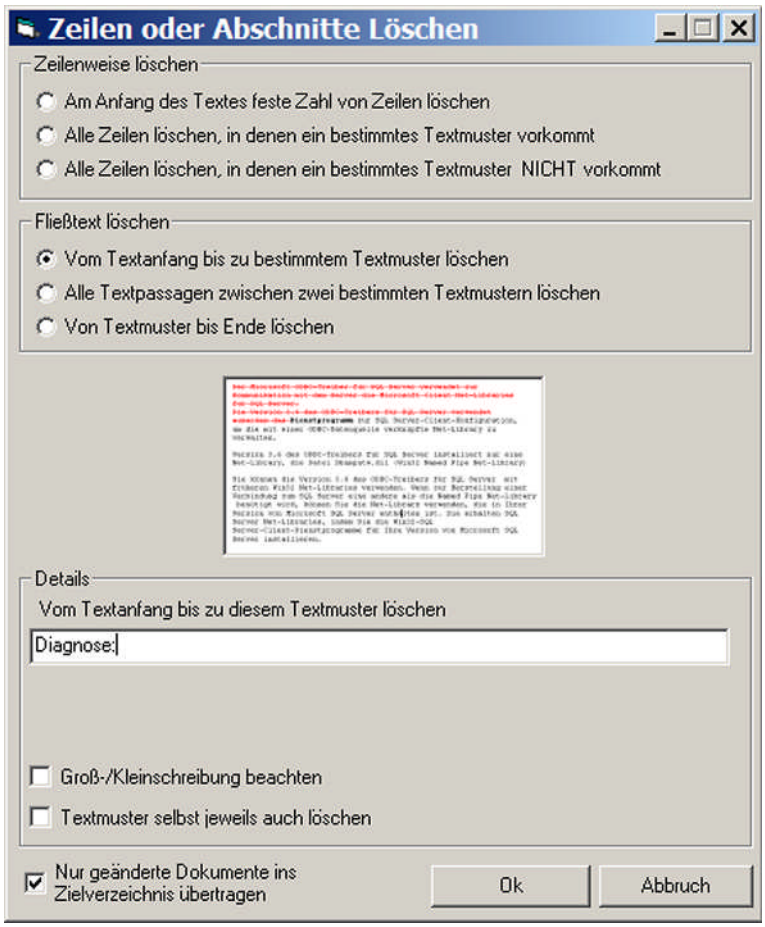
Durch einen Klick auf **Ok** werden alle Dateien des Quellverzeichnisses gemäß Anweisung bearbeitet und in das Zielverzeichnis gespeichert.

c) Löschen aller Zeilen, in denen ein bestimmter Suchbegriff nicht vorkommt.

Diese Option ist im engen Zusammenhang mit dem Tool REMOVERR zu sehen und wird dort beschrieben.

d) Löschen des Textes vom Textanfang bis zu einem bestimmten Suchbegriff.

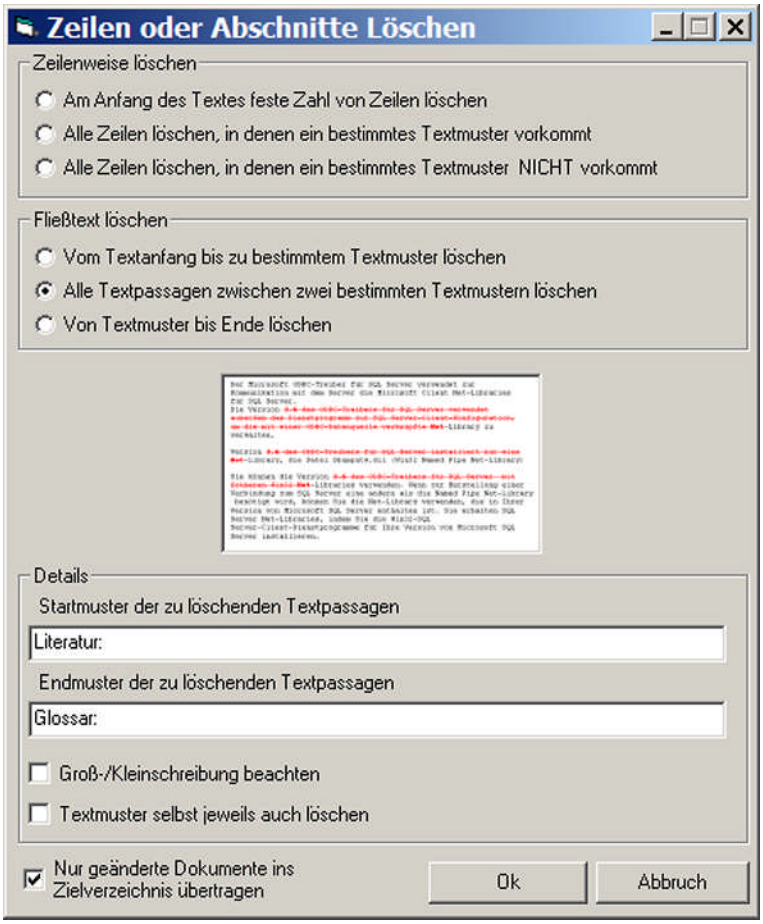
Im vorliegenden Beispiel wird jeder Text bis zu dem Wort „Diagnose“ gelöscht. Dies ist dann angezeigt, wenn Briefkopf oder Adressbereich variabel lang sind.



Ist im Dokument das Wort (hier „Diagnose:) nicht enthalten, wird kein Text gelöscht.

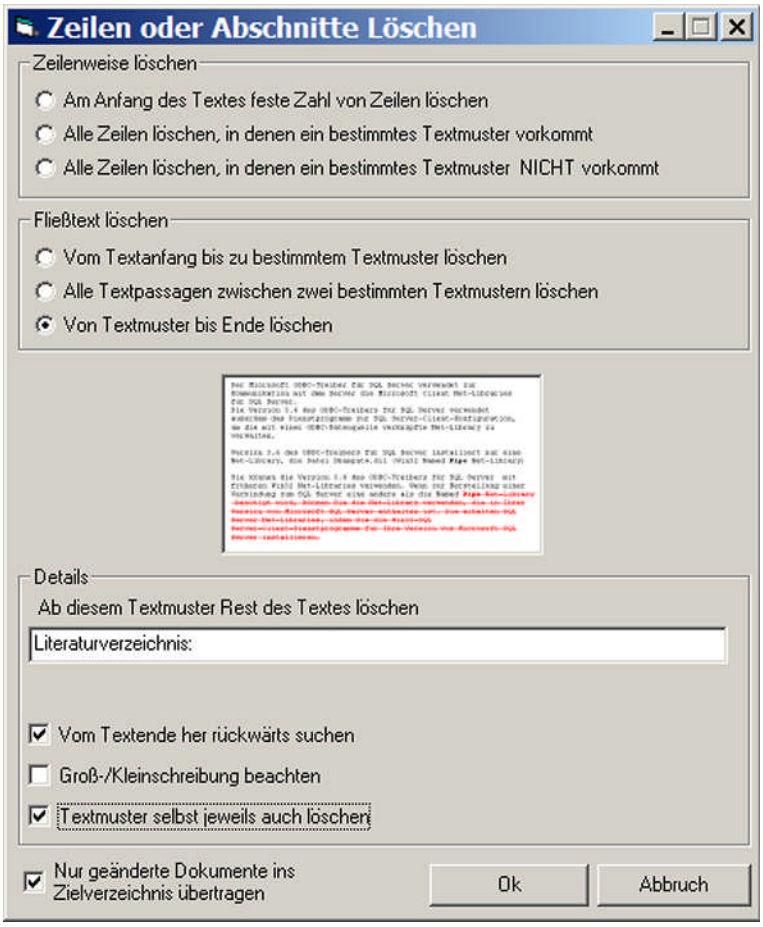
Man kann zusätzlich angeben, ob der Suchbegriff selbst auch gelöscht werden soll („Textmuster selbst jeweils auch löschen“).

d) Es werden alle Textabschnitte zwischen zwei Suchbegriffen gelöscht.



Im vorliegenden Beispiel wird jede Textpassage gelöscht, die sich zwischen den Suchbegriffen „Literatur:“ und „Glossar:“ befindet. Besonders bei Veröffentlichungen kann man Kapitel mit störenden Namen und Literaturstellen leicht aus im Quelltext entfernen.

e) Löschen von Textpassagen ab einem Suchbegriff bis zum Dokumentende.



Hier wird sämtlicher Text ab dem Suchbegriff „Literaturverzeichnis:“ bis zum Textende gelöscht.

Bei dieser Funktion ist es wichtig, dass man die Möglichkeit hat, den Suchbegriff vom Textende her rückwärts zu suchen. Es wird im Text sicher öfters auf den Suchbegriff verwiesen. Gelöscht werden soll aber meistens nur vom letzten Auftreten des Suchbegriffs bis zum Textende.
Im vorliegenden Beispiel soll das Kapitel „Literaturverzeichnis:“ gelöscht werden, das sich am Textende befindet.

Bei allen Funktionen von DELLINES hat man die Option, ob alle Dokumente in das Zielverzeichnis übertragen werden sollen, oder nur diejenigen, bei denen eine oder mehrere Zeilen gelöscht wurden.

B) Die Erzeugung von Wortlisten

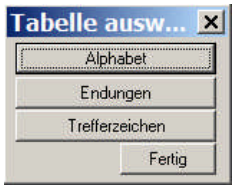
Bei der Verwendung des NaturalVocTools von Nuance ist ein wichtiger Prozessschritt, dem NaturalVocTool sog. Wortlisten zur Verfügung zu stellen.

Das ToolSet TextPrep generiert, nachdem der Quelltext mit TextPrep bearbeitet, anonymisiert und korrigiert wurde, eine Reihe sehr sinnvoller Wortlisten.

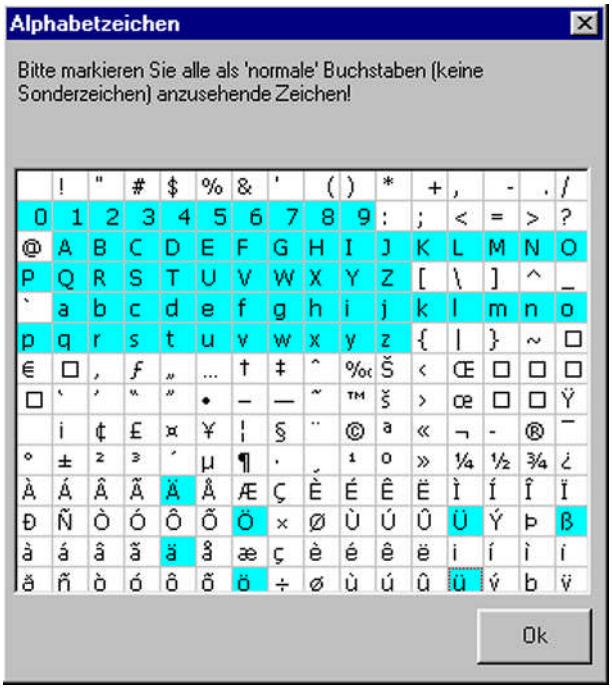
8. ABBREVNS

Dieses Tool erzeugt eine Liste aller Abkürzungen, die in den Dateien im Quellverzeichnis vorkommen. Diese Abkürzungen werden in die Datei „Abkürzungen.txt“ im Wortlistenverzeichnis eingetragen. In einem separaten Schritt ist dann die Liste der Abkürzungen mit der gesprochenen Form zu ergänzen.

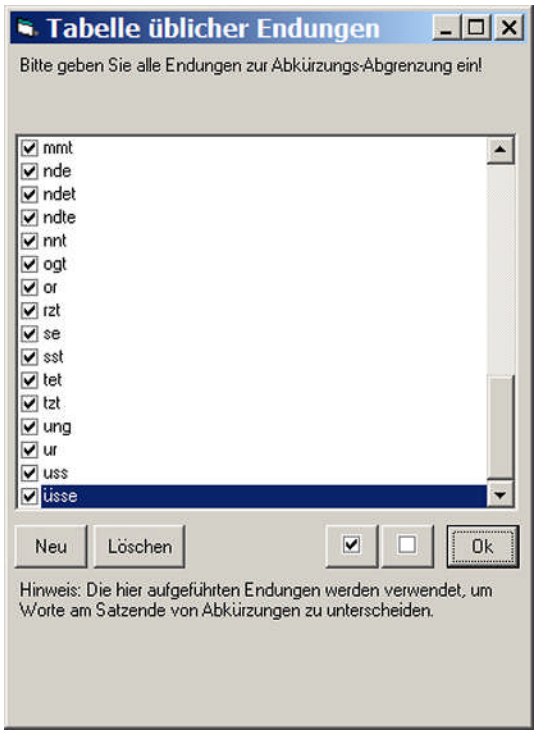
Nachdem man im Hauptmenu das Tool ABBREVNS markiert hat, klickt man (beim erstmaligen Aufruf dieses Tools) auf die Taste „Zugehörige Tabelle bearbeiten“.



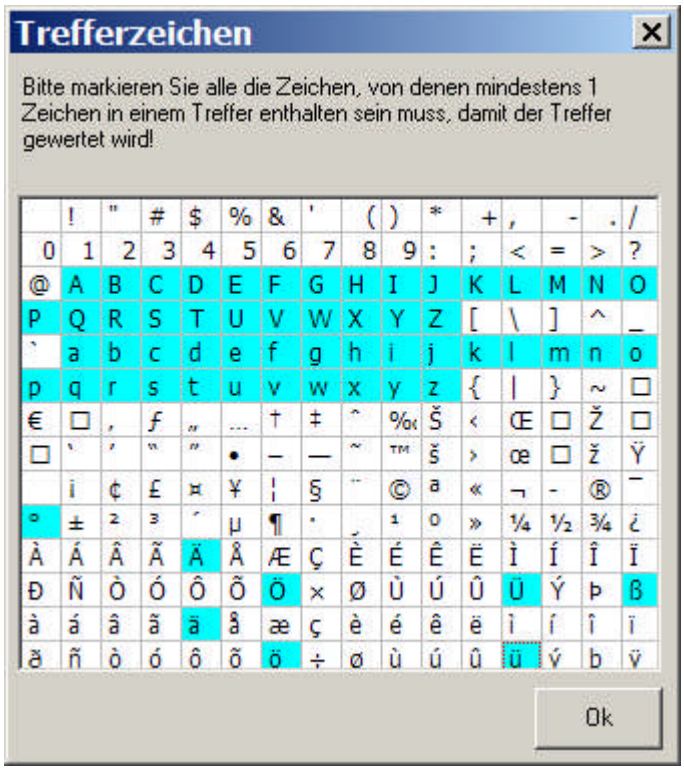
Bei **Alphabet** wird eine Tabelle aller Zeichen angezeigt, die festlegt, welche Zeichen zum Alphabet gehören.



Bei **Endungen** erscheint eine Liste, in der man Wortenden angeben kann, die typischerweise keine Abkürzungen sind:

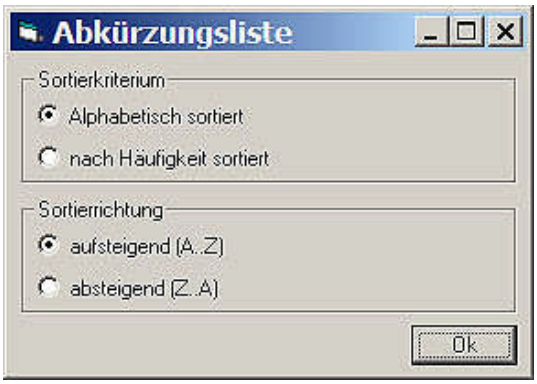


In der Tabelle **Trefferzeichen** gibt man an, von welchem Zeichensatz mindestens 1 Zeichen vorkommen muss, um als Treffer zu gelten. Auf diese Weise kann man erreichen, dass Gebilde, die nur aus Ziffern und Sonderzeichen bestehen (z.B. Kommazahlen, Datum) als Treffer gewertet werden.



Ist die Liste der Tabellen fertig gestellt, klickt man auf die Taste **Fertig**.

Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool ABBREVNS aus und klickt auf **Ok.** Im nächsten Schritt kann ausgewählt werden, ob die Abkürzungen alphabetisch oder nach Häufigkeit sortiert sein sollen:

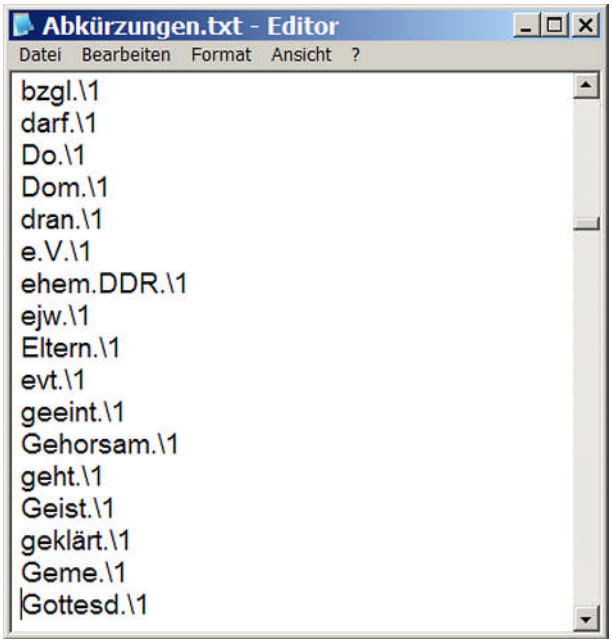


„Nach Häufigkeit“ ist dann sinnvoll, wenn man Abkürzungen mit geringer Häufigkeit in 1 Arbeitsgang entfernen möchte. Ansonsten dürfte „alphabetisch aufsteigend“ die übliche Sortierfolge sein.

Nach Klick auf OK führt das Tool die Analyse durch und endet sinngemäß mit der Meldung:

Das Tool hat nun alle Wörter gefunden, die mit einem Punkt enden und danach ein klein geschriebenes Wort folgt, vermindert um alle Wörter, deren Ende in der Liste **Endungen** (s.o.) aufgeführt sind.

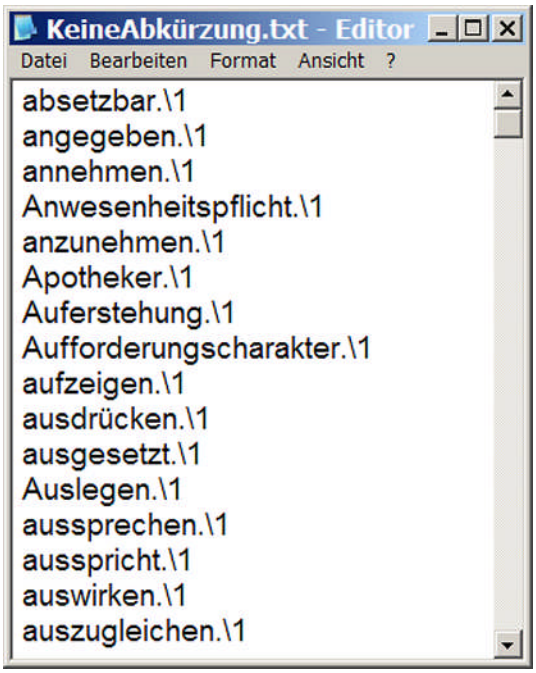
Das Ergebnis (Datei „Abkürzungen.txt“ im Wortlistenverzeichnis) könnte wie folgt aussehen:



Die Zahl am Ende gibt die Anzahl der Treffer im Gesamttext an. Dies kann als Hilfe dienen, wie wichtig diese Abkürzung ist. Mit dem Tool **MODFYLST** lassen sich gezielt die Wörter eliminieren, die z.B. nur einmal vorkommen.

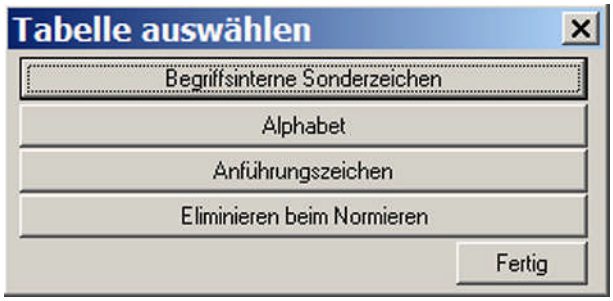
Nun sind die Abkürzungen manuell zu bearbeiten, d.h. Scheinabkürzungen und seltene Abkürzungen sind zu entfernen. Bei den verbleibenden Abkürzungen ist die gesprochene Form mit dem Tool **2LISTS** hinzuzufügen, wenn schon ein größerer Bestand an Abkürzungen incl. der gesprochenen Form existiert. Eine andere Alternative ist das Tool **MODFYLST** (s. dort).

Zusätzlich zu dieser Liste wurde auch eine Liste „KeineAbkürzung“ (im Wortlistenverzeichnis) erzeugt, die alle eliminierten Wörter enthält. Diese sollte man kurz kontrollieren, ob eventuell doch eine echte Abkürzung fälschlicherweise entfernt wurde:



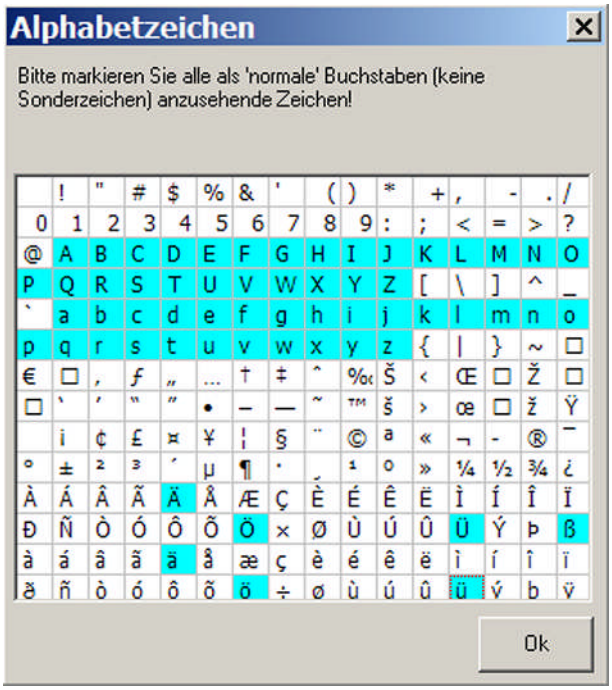
9. INITIALS

Das Tool INITIALS erzeugt eine Wortliste, die alle Begriffe enthält, die einen Großbuchstaben (auch) an anderer als der ersten Stelle aufweisen
Vor Ausführung des Tools müssen (in der Regel nur bei der erstmaligen Benutzung) die zugehörigen Tabellen bearbeitet werden:

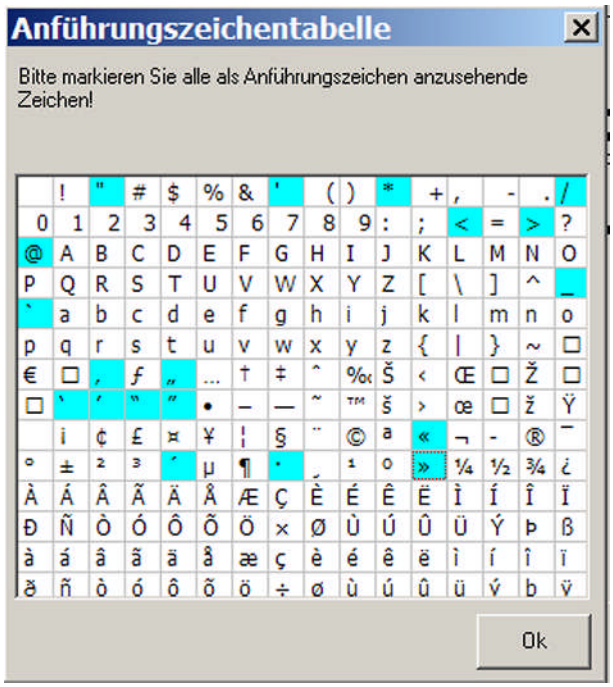


Die Tabelle **Begriffsinterne Sonderzeichen** ist für dieses Tool ohne Bedeutung.

Die Tabelle **Alphabet**
Dort sind alle Zeichen blau markiert (schattiert), die in den zu suchenden Begriffen vorkommen dürfen.



Die Tabelle **Anführungszeichen**
In dieser Tabelle sind alle Zeichen markiert, die als Anführungszeichen anzusehen sind. Sie werden vor der Ausführung des Tools entfernt, da sie kein Bestandteil der zu suchenden Begriffe sind.

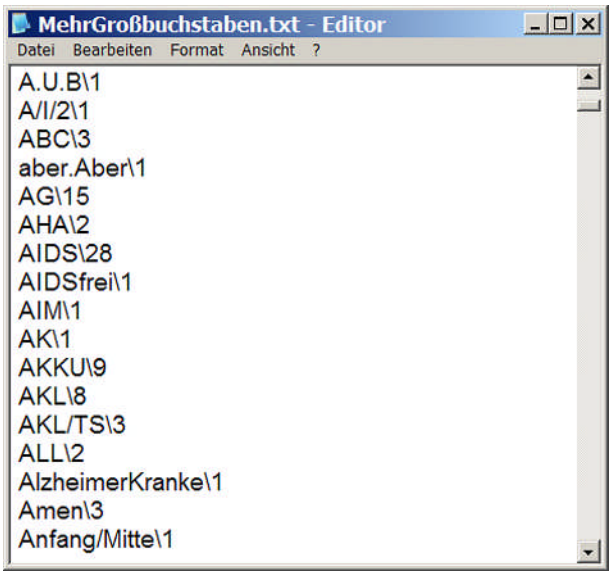


Die Tabelle **Eliminieren beim Normieren**

Bevor der Text nach Ausdrücken wie „ADAC“ oder „eMail“ durchsucht wird, wird der Text normiert. D.h., alle überflüssigen Zeichen wie z.B., Klammern, Hochkommas, usw. werden durch ein Leerzeichen ersetzt. Der Abstand der einzelnen Wörter wird auf 1 Leerzeichen reduziert. Die Tabelle „Eliminieren beim Normieren“ kennzeichnet alle Zeichen, die durch ein Leerzeichen ersetzt werden.

Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool INITIALS aus und klickt auf **Ok.** Das Tool führt die Analyse durch und endet sinngemäß mit der Meldung:

Das Ergebnis (Datei „MehrGroßbuchstaben.txt“ im Wortlistenverzeichnis) könnte wie folgt aussehen:



Die Zahlen hinter den Begriffen geben die Anzahl der Treffer im Gesamttext an. Man kann auf dieser Basis entscheiden, ob der Begriff in das Vokabular aufgenommen werden soll oder nicht. Mit dem Tool **MODFYLST** lassen sich gezielt die Wörter eliminieren, die z.B. nur einmal vorkommen.

Nun sind diese Begriffe manuell zu bearbeiten, d.h. ungültige Begriffe sind zu entfernen und bei den verbleibenden Begriffen ist die gesprochene Form mit dem Tool **2LISTS** hinzuzufügen, wenn schon ein größerer Bestand an Abkürzungen incl. der gesprochenen Form existiert. Eine Alternative ist der Einsatz des Tools **MODFYLST** (s. dort).

10. NOVOWELS

Mit diesem Tool werden alle Begriffe gefunden, die keine Vokale enthalten. Damit sollen möglichst viele Maßeinheiten gefunden werden. Beispiele: mg/ml, µg, °C, qm. Diese Begriffe werden in die Datei „VokalfreieBegriffe.txt“ im Wortlistenverzeichnis eingetragen. In einem separaten Schritt ist dann die Liste der Begriffe mit der gesprochenen Form zu ergänzen.

Vor der (erstmaligen) Ausführung des Tools sind folgende Tabellen zu überprüfen:

Die Tabelle **Vokale**
Die hier markierten Zeichen werden als Vokale angesehen und sind in den zu findenden Begriffen nicht enthalten.

Die Tabelle **Eliminieren beim Normieren**

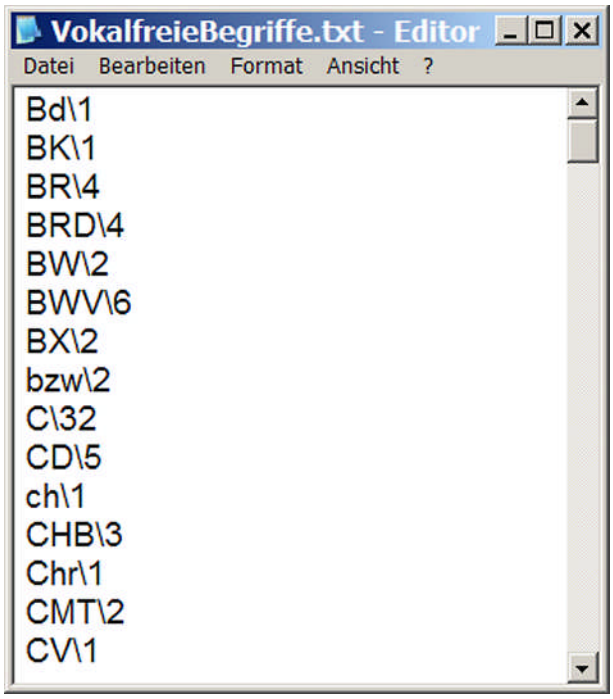
Bevor der Text nach Ausdrücken wie „mg/dl“ oder „°C“ durchsucht wird, wird der Text normiert. D.h., alle überflüssigen Zeichen wie z.B., Klammern, Hochkommas, usw. werden entfernt, ebenso überflüssige Leerzeichen. Der Abstand der einzelnen Wörter wird auf 1 Leerzeichen reduziert. Die Tabelle „Eliminieren beim Normieren“ kennzeichnet alle Zeichen, die eliminiert bzw. durch ein Leerzeichen ersetzt werden.

Achtung: Markieren Sie in dieser Tabelle keine Satzzeichen! Sonst werden diese durch ein Leerzeichen ersetzt und es gehen für die Bigrammstatistiken des Dragon NaturalVocTools wertvolle Informationen verloren.

In der Tabelle **Trefferzeichen** gibt man an, von welchem Zeichensatz mindestens 1 Zeichen vorkommen muss, um als Treffer zu gelten. Auf diese Weise kann man erreichen, dass Gebilde, die nur aus Ziffern und Sonderzeichen bestehen (z.B. Kommazahlen, Datum) als Treffer gewertet werden.

Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool INITIALS aus und klickt auf **Ok.** Nach Fertigstellung meldet sich das Tool sinngemäß mit folgender Nachricht:

Das Ergebnis (Datei „VokalfreiBegriffe.txt“ im Wortlistenverzeichnis) könnte wie folgt aussehen:



Die Zahlen hinter den Begriffen geben die Anzahl der Treffer im Gesamttext an. Man kann auf dieser Basis entscheiden, ob der Begriff in das Vokabular aufgenommen werden soll oder nicht. Mit dem Tool **MODFYLST** lassen sich gezielt die Wörter eliminieren, die z.B. nur einmal vorkommen.

Nun sind diese Begriffe manuell zu bearbeiten, d.h. ungültige Begriffe sind zu entfernen und bei den verbleibenden Begriffen ist die gesprochene Form mit dem Tool **2LISTS** hinzuzufügen, wenn schon ein größerer Bestand an Abkürzungen incl. der gesprochenen Form existiert. Eine Alternative ist das Tool **MODFYLST** (s. dort).

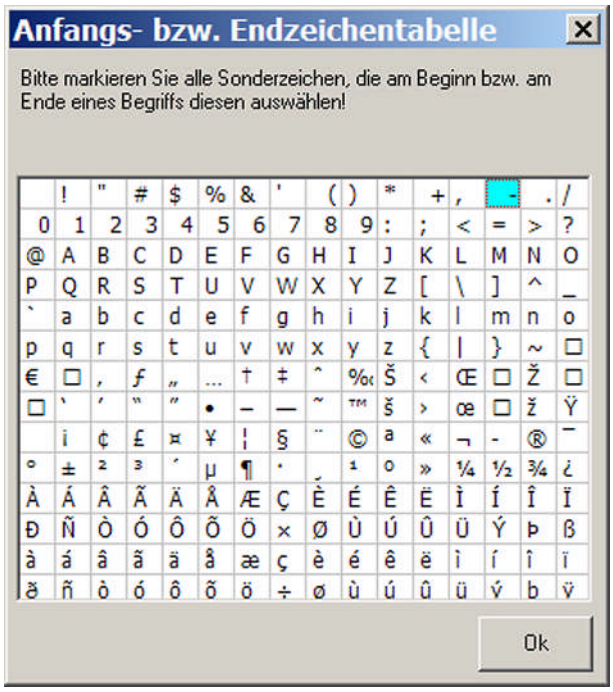
Hinweis:
Die Ergebnislisten der Tools INITIALS und NOVOWELS überlappen sich. Es ist daher sinnvoll, diese beiden Ergebnislisten zusammenzuführen, zu sortieren, Duplikate zu entfernen und dann zu bearbeiten. Das Sortieren und Duplikate entfernen erreicht man am leichtesten, indem man diese Datei im Tool **2LISTS** (ohne eine zweite Liste) aufruft und speichert. Das Tool 2LISTS führt diese Arbeitsschritte automatisch durch.

11. PHRASES

Um eine hohe Erkennungsrate zu erzielen, ist es für ein Fachvokabular sinnvoll, auch Begriffe hinzuzufügen, die durch Wortgebilde wie z.B. „Hals- und Beinbruch“ oder „Knochenfraktur und –behandlung“ entstehen. Das Tool PHRASES extrahiert alle Begriffe, die z.B. einen Ergänzungsstrich enthalten. Diese Begriffe werden in die Datei „AnfangsZeichenBegriffe.txt“ im Wortlistenverzeichnis eingetragen.

Vor Ausführung des Tools sind mehrere Tabellen zu bearbeiten.
Man klickt daher auf „Zugehörige Tabellen bearbeiten:

In der Tabelle **Anfangs- bzw. Endzeichen** wird definiert, welche Zeichen als Trennzeichen (Ergänzungsstrich) aufzufassen sind.

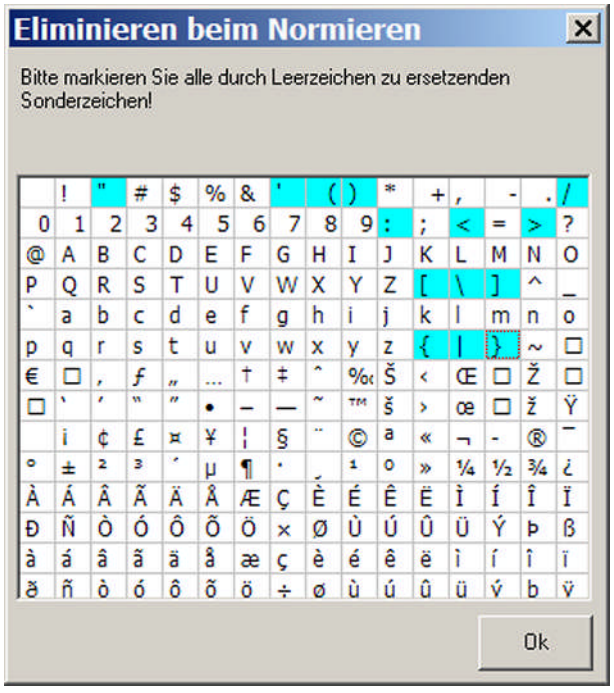


Üblicherweise selektiert man den Bindestrich. Doch die Suche kann universeller eingesetzt werden: Sucht man z.B. alle Copyright-Begriffe, müsste man nur das Zeichen © auswählen und schon bekäme man alle Namen, denen ein Copyright-Zeichen vorangesetzt bzw. nachgestellt ist.

In der Tabelle **Combiworte** definiert man alle Worte, die zwischen den beiden Inhaltsworten stehen. Im Falle „Hals- und Beinbruch“ das Wort „und“.

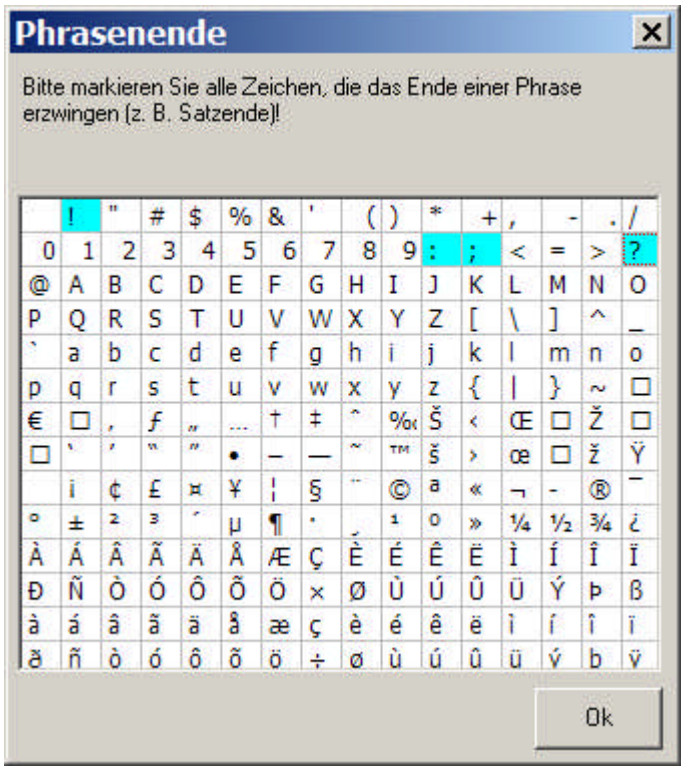
Eine Liste von sog. Combiworten, d.h. Verbindungsworten, könnte wie folgt aussehen:

Bevor der Text nach Ausdrücken wie „Hals- und Beinbruch“ durchsucht wird, wird der Text normiert. D.h., alle überflüssigen Zeichen wie z.B., Klammern, Hochkommas, usw. werden entfernt, ebenso überflüssige Leerzeichen. Der Abstand der einzelnen Wörter wird auf 1 Leerzeichen reduziert. Die Tabelle **Eliminieren beim Normieren** kennzeichnet alle Zeichen, die eliminiert bzw. durch ein Leerzeichen ersetzt werden.

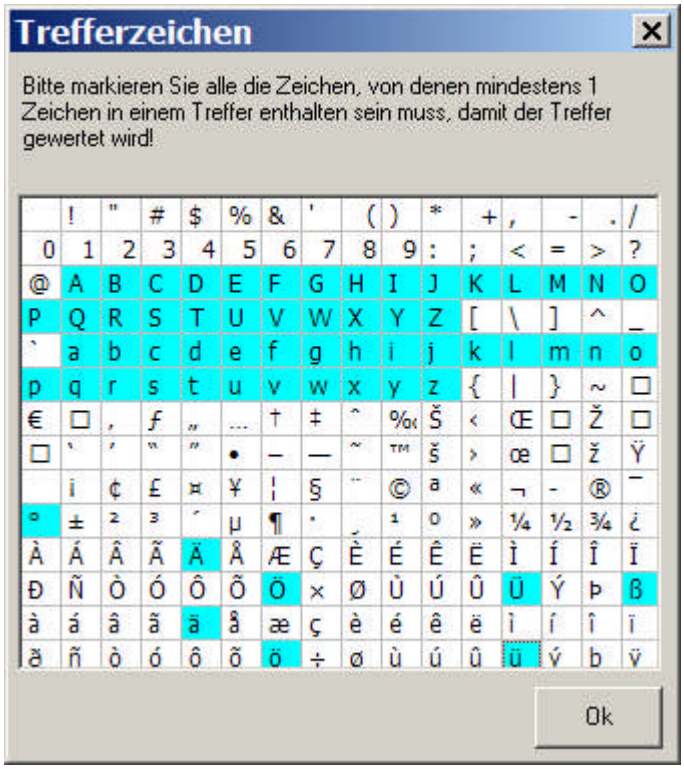


Achtung: Markieren Sie in dieser Tabelle keine Satzzeichen! Sonst gehen für die Bigrammstatistiken des Dragon NaturalVocTools wertvolle Informationen verloren.

In der Tabelle **Phrasenende** gibt man an, welche Sonderzeichen ein Ende der Phrase erzwingen, d.h. nie Bestandteil einer gesuchten Phrase sein sollen.

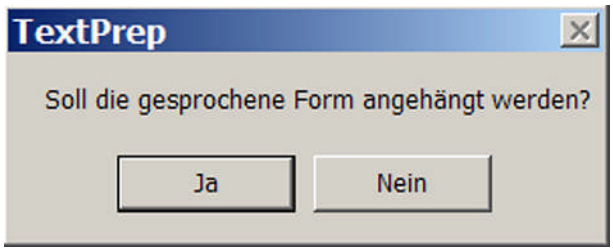


In der Tabelle **Trefferzeichen** gibt man an, von welchem Zeichensatz mindestens 1 Zeichen vorkommen muss, um als Treffer zu gelten. Auf diese Weise kann man erreichen, dass Gebilde, die nur aus Ziffern und Sonderzeichen bestehen (z.B. Kommazahlen, Datum) als Treffer gewertet werden.

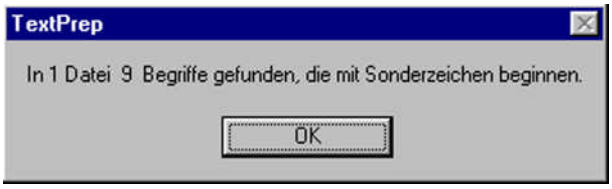


Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool PHRASES aus und klickt auf **Ok.**

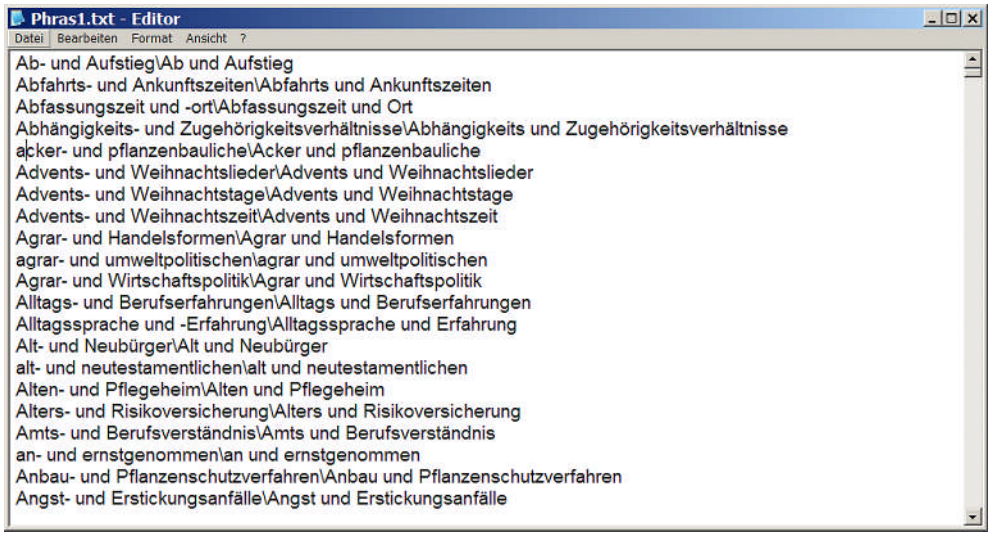
Man wird gefragt, ob die gesprochene Form (automatisch) generiert werden soll:



Das Tool führt die Analyse durch und endet sinngemäß mit der Meldung:



Das Ergebnis (Datei „Phrasen.txt“ im Wortlistenverzeichnis) könnte wie folgt aussehen:



In diesem Fall wurde sinnvollerweise die gesprochene Form automatisch hinzugefügt.

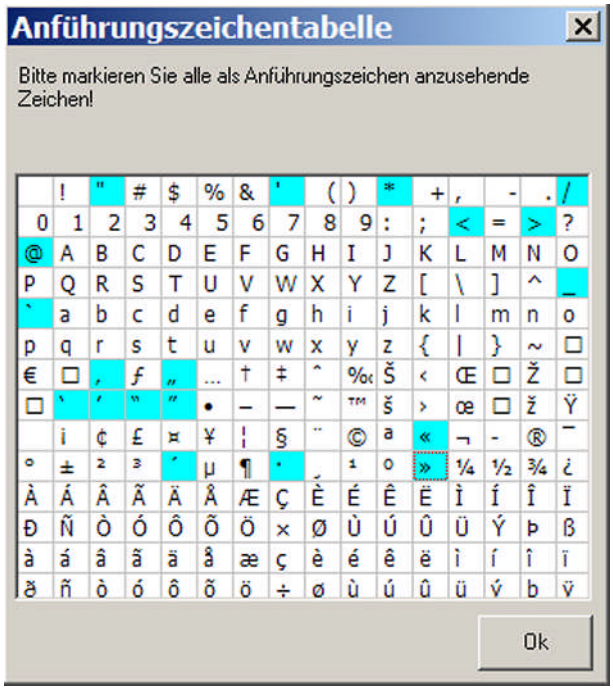
12. MULTTERM

Sehr wichtig für Fachvokabulare sind Mehrwortbegriffe wie z.B. „Mutter-Kind-Beziehung“. **Es soll dem Diktierenden erspart bleiben, die Bindestriche mitdiktieren zu müssen**. Es ist daher sehr hilfreich, diese Mehrwortbegriffe aufzufinden und in einer Wortliste mit der gesprochenen Form (ohne Bindestriche) bereitzustellen.
MULTTERM findet diese Begriffe, ergänzt sie auf Wunsch mit der gesprochenen Form, und speichert sie in der Datei „Mehrwortbegriffe“ in das Wortlistenverzeichnis. Es kann festgelegt werden, ob auch Mehrwortbegriffe gefunden werden sollen, die ein anderes Sonderzeichen als den Bindestrich enthalten. So werden z.B. mit dem Zeichen „@“ alle E-Mail-Adressen gefunden. Weitere mögliche Listen: Schrägstrich-Begriffe, Hochkomma-Wörter, Punkt-Wörter (z.B. ICD10-Codes), Begriffe mit Prozentzeichen usw.

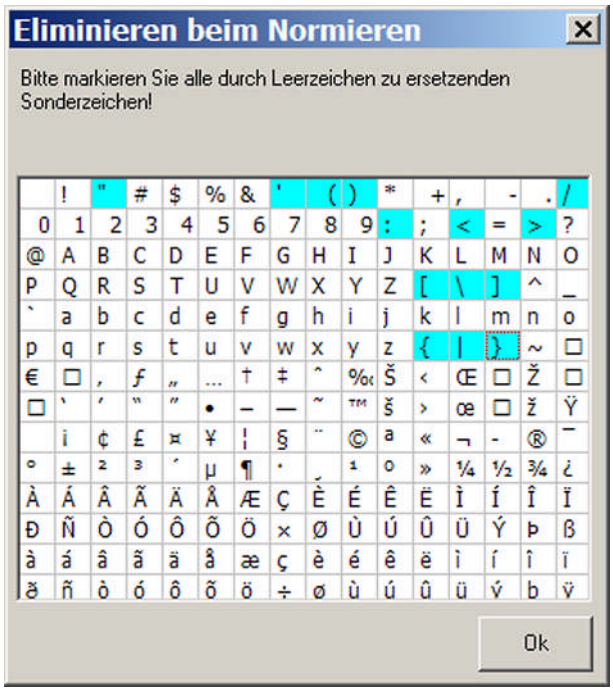
Nachdem man im Hauptmenu MULTTERM markiert hat, klickt man (beim erstmaligen Aufruf dieses Tools) auf die Taste „Zugehörige Tabelle bearbeiten“. Daraufhin wird eine Tabelle aller Zeichen angezeigt, die festlegt, welche Zeichen als begriffsinterne Sonderzeichen anzusehen sind.

Üblicherweise werden neben dem Bindestrich (**-**) auch das Auslassungszeichen (**’**) oder der Schrägstrich (***/***) selektiert. Mit einem Klick auf **Ok** wird die Tabelle zurückgeschrieben.

Die Tabelle **Anführungszeichen**
In dieser Tabelle sind alle Zeichen markiert, die als Anführungszeichen anzusehen sind. Sie werden vor der Ausführung des Tools entfernt, da sie kein Bestandteil der zu suchenden Begriffe sind.



Die Tabelle **Eliminieren beim Normieren**
Bevor der Text nach Ausdrücken wie „ADAC“ oder „eMail“ durchsucht wird, wird der Text normiert. D.h., alle überflüssigen Zeichen wie z.B., Klammern, Hochkommas, usw. werden entfernt, ebenso überflüssige Leerzeichen. Der Abstand der einzelnen Wörter wird auf 1 Leerzeichen reduziert. Die Tabelle „Eliminieren beim Normieren“ kennzeichnet alle Zeichen, die eliminiert bzw. durch ein Leerzeichen ersetzt werden.



Achtung: Markieren Sie in dieser Tabelle keine Satzzeichen! Sonst werden diese durch ein Leerzeichen ersetzt und es gehen für die Bigrammstatistiken des Dragon NaturalVocTools wertvolle Informationen verloren.

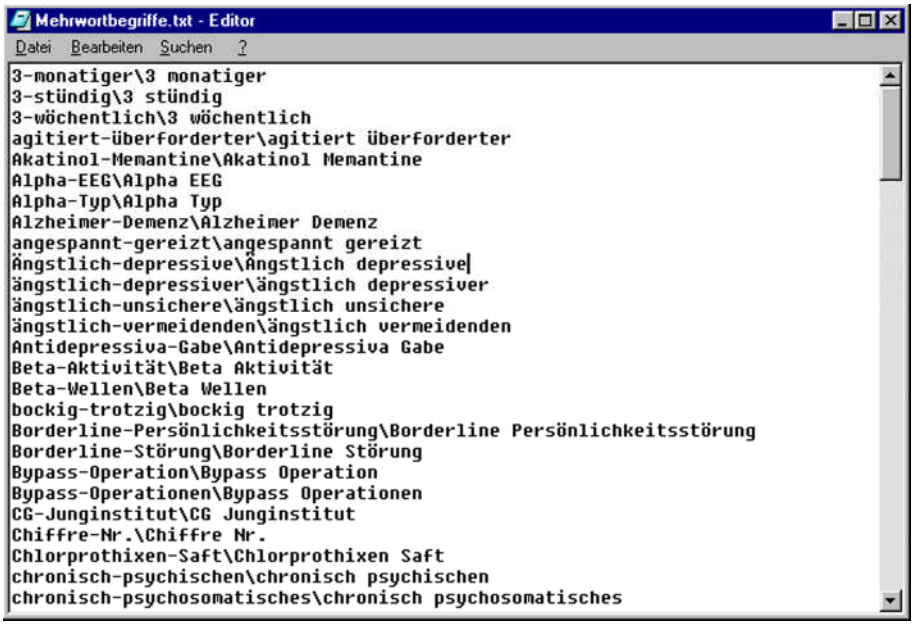
In der Tabelle **Trefferzeichen** gibt man an, von welchem Zeichensatz mindestens 1 Zeichen vorkommen muss, um als Treffer zu gelten. Auf diese Weise kann man erreichen, dass Gebilde, die nur aus Ziffern und Sonderzeichen bestehen (z.B. Kommazahlen, Datum) als Treffer gewertet werden.

Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool MULTTERM aus und klickt auf **Ok.** Man wird gefragt, ob die gesprochene Form automatisch generiert werden soll:

Dies wird mit einem Klick auf **Ja** oder **Nein** beantwortet. Das Tool führt die Begriffssuche durch, speichert das Ergebnis in der Datei „Mehrwortbegriffe.txt“ in die Stammdatei und endet mit folgender sinngemäßer Meldung:

Mit einem Klick auf **OK** schließt sich dieses Fenster.

Das Ergebnis (Datei „Mehrwortbegriffe.txt“ im Stammverzeichnis) könnte wie folgt aussehen:



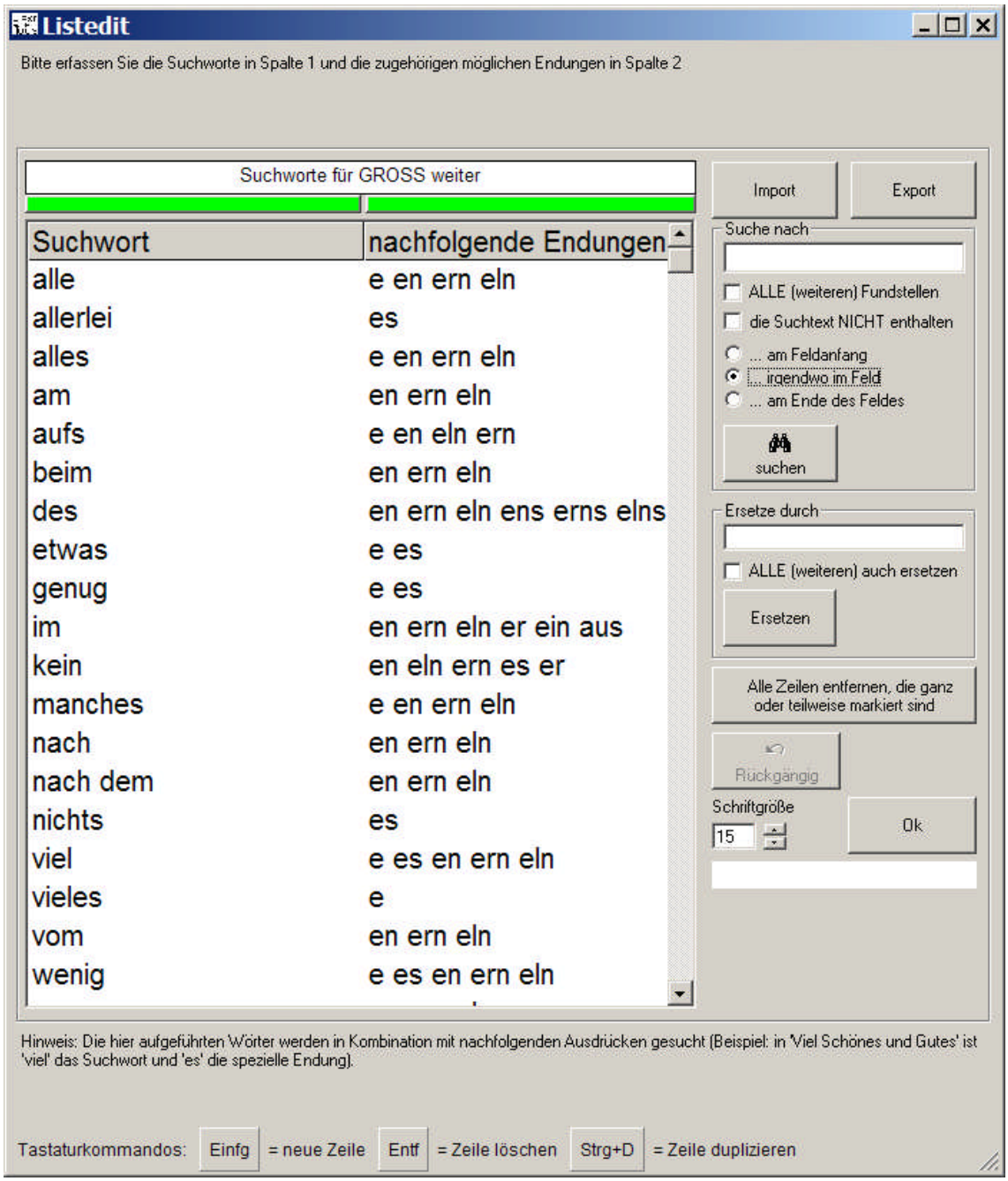
Die Ergebnisliste muss nur noch geringfügig nachgearbeitet werden, da evtl. enthaltene Abkürzungen in der gesprochenen Form nicht dem automatisch generierten Vorschlag entsprechen. Zum Beispiel muss „Nr.“ als gesprochene Form durch „Nummer“ ersetzt werden. Generell ist das Nacharbeiten mit dem Tool **MODFYLST** zu empfehlen.

13. NEXTWORD

Adjektive und Verben werden oft im Zusammenhang mit den Wörtern "alles, etwas, nichts, …usw. als Substantive verwendet. Diese werden jedoch bei Dragon NaturallySpeaking oft klein geschrieben. Dies kann verhindert werden, wenn man diese Wörter als Begriff in das Vokabular aufnimmt, so z.B. "alles Schöne, nichts Gutes, etwas Erfreuliches, …usw.).

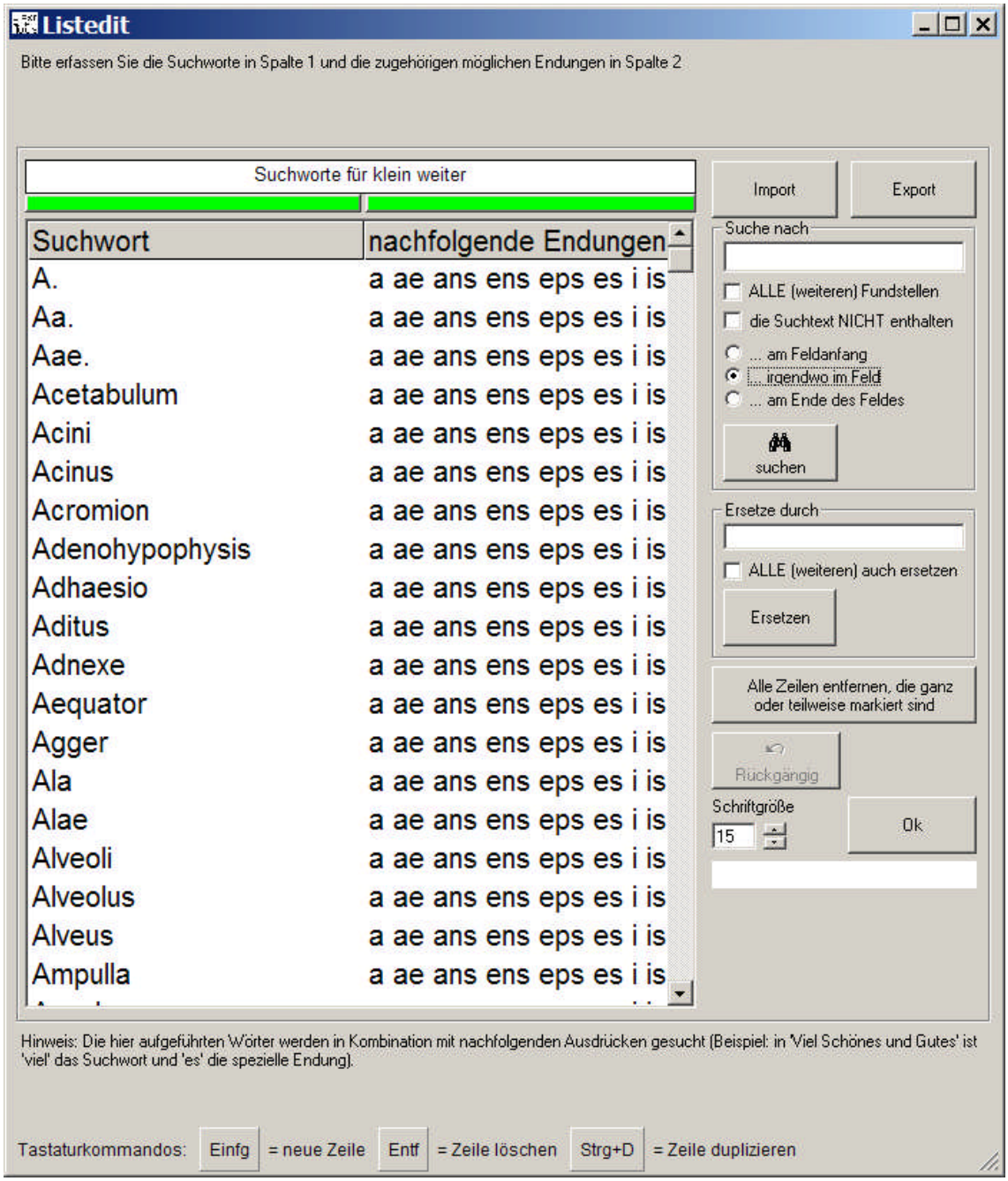
Umgekehrt möchte man auch Mehrwort-Ausdrücke wie z.B. "Diabetes mellitus" erfassen. NEXTWORD findet auch diese Begriffe.

Vor der Ausführung des Tools müssen auch hier die zugehörigen Tabellen bearbeitet werden: Liste **Suchworte für GROSS weiter,** Liste **Suchworte für klein weiter** und Liste **Füllworte**:



Die Tabelle **Suchworte für GROSS weiter** enthält alle Wörter, nach denen ein substantiviertes Adjektiv bzw. Verb stehen könnte. Um die Liste der gewünschten Treffer einzuengen, gibt man für jedes Suchwort an, welche Endungen der Folgewörter erlaubt sind. Dies filtert viele unsinnige "Treffer" aus.

Die Tabelle **Suchworte für klein weiter** enthält alle Wörter, nach denen alle weiteren klein geschriebenen Wörter miterfasst werden sollen, die ganz bestimmte Endungen aufweisen. Auf diese Weise lassen sich medizinische Ausdrücke sehr gut extrahieren, um sie als separate Begriffe in das Vokabular einzubringen.



Die in der Tabelle **Füllworte** enthaltenen Wörter werden ungeachtet der Suchregel als zum gesuchten Ausdruck zugehörig betrachtet. Beispiel für das Füllwort "und": Es wird auch der Begriff "alles Schöne und Gute" gefunden.

Zum Start des Tools wählt man es im Hauptmenu aus und klickt dann auf Ok.
Das Tool schreibt 2 Ergebnislisten in das Wortlistenverzeichnis:

Das Ergebnis könnte wie folgt aussehen:

BegriffeGrossWeiter.txt

---

beim Abtrocknen
beim Klettern
im Einzelnen
im Freien
nach Öffnen
nach Verlassen
zum Auftragen
zum Reinigen

BegriffeKleinWeiter.txt

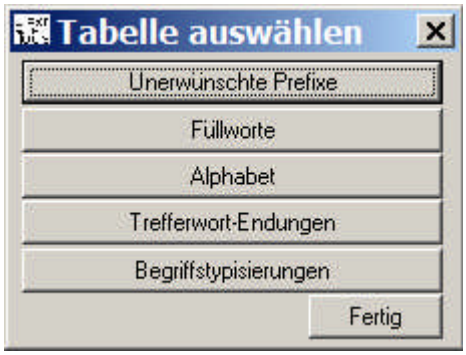
A. dorsalis pedis
Arcus tendineus
Caput radiale
Epicondylus humeri lateralis
Ligamentum collaterale
M. extensor pollicis brevis
N. suralis
Ramus interosseus dorsalis

14. XPRESSNS

Es gibt im Deutschen oft Namen von Organisationen usw., die aus mehreren großgeschriebenen Wörtern bestehen, ggf. auch mit kleingeschriebenen Füllwörtern. Eigenschaftswörter, die meist am Anfang solcher Eigennamen stehen, werden hierbei auch groß geschrieben.

Beispiele: "Deutsches Rotes Kreuz", "Internationales Institut für Weltraumforschung". XPRESSNS erzeugt aus den Quelltexten eine Wortliste solcher Ausdrücke.

Vor der Ausführung dieses Tools müssen noch 5 Tabellen bearbeitet werden:



Um viele Scheintreffer zu vermeiden, vor allem durch großgeschriebene Wörter am Satzanfang, werden in der Tabelle **Unerwünschte Prefixe** diese Wörter aufgelistet:

Auch werden sog. Füllwörter zugelassen, und als zum gesuchten Mehrwort-Eigennamen zugehörig betrachtet:

Die Tabelle **Alphabet** wurde schon in anderen Kapiteln beschrieben.

Als „Trefferwörter“ werden im Tool XPRESSNS Eigenschaftswörter angesehen, die am Anfang des gesuchten Mehrwort-Eigennamens stehen. In der Tabelle **Trefferwort-Endungen** sind daher alle üblichen Endungen von Eigenschaftswörtern aufzulisten. Die Tabelle könnte daher so aussehen:

Um das Suchergebnis weiter einzugrenzen, wird in der Tabelle **Begriffstypisierungen** festgelegt, welchen Aufbau die gesuchten Mehrwort-Eigennamen haben sollen.

Dabei ist:

T = Trefferwort , d.h. ein groß geschriebenes Eigenschaftswort mit einer der Endungen, wie sie in der Tabelle „Trefferwort-Endungen“ spezifiziert wurden.

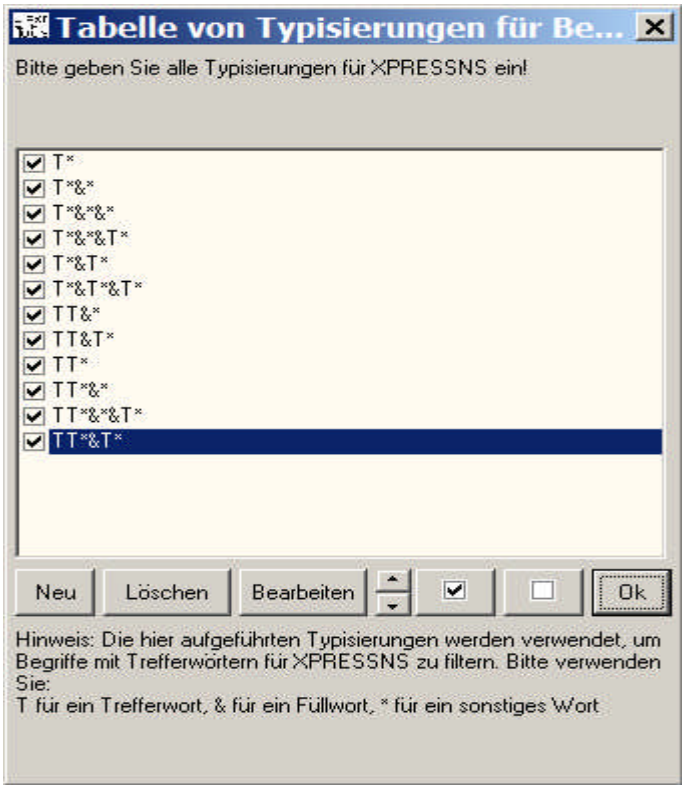
& = Füllwort (s. Tabelle „Füllworte“)

* = sonstiges groß geschriebenes Wort

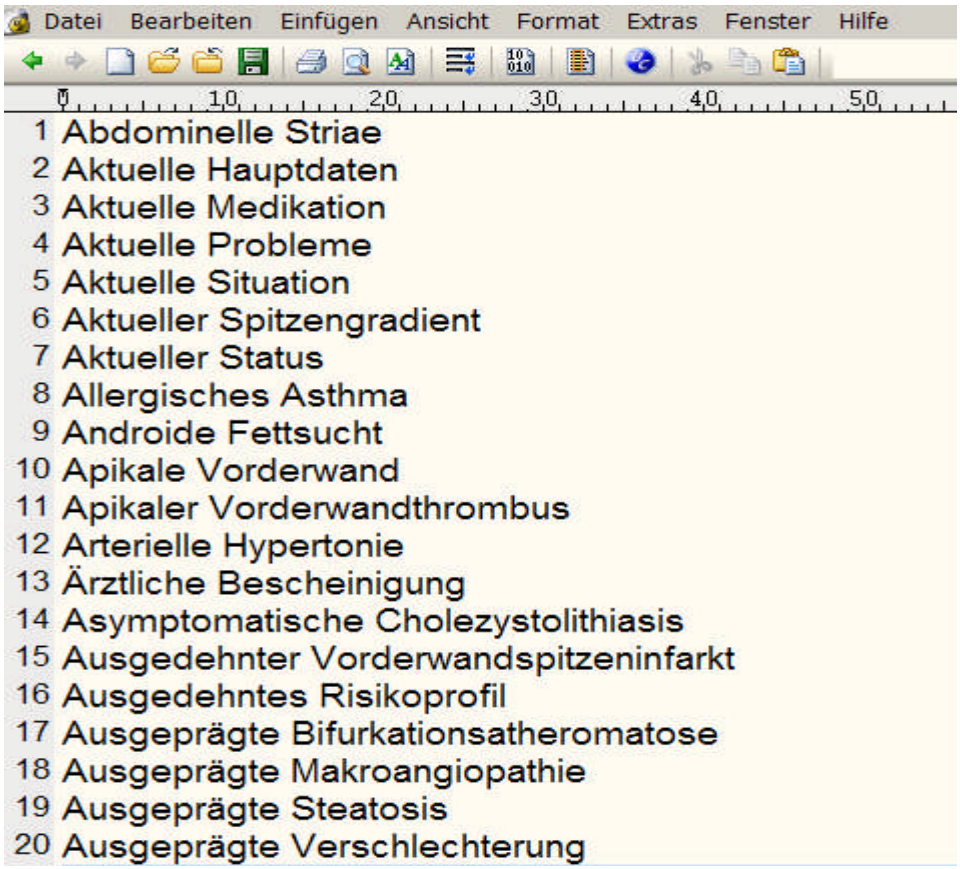
Beispiele:

Arterielle Hypertonie	= T*
Deutsches Rotes Kreuz	= TT*
Deutsches Institut für Moderne Kunst	= T*&T*

Alle gewünschten Typisierungen sind in der Tabelle **Begriffstypisierungen** einzutragen. Dies könnte so aussehen:



Zum Start des Tools wählt man es im Hauptmenu aus und klickt dann auf Ok.
Das Ergebnis könnte wie folgt aussehen:



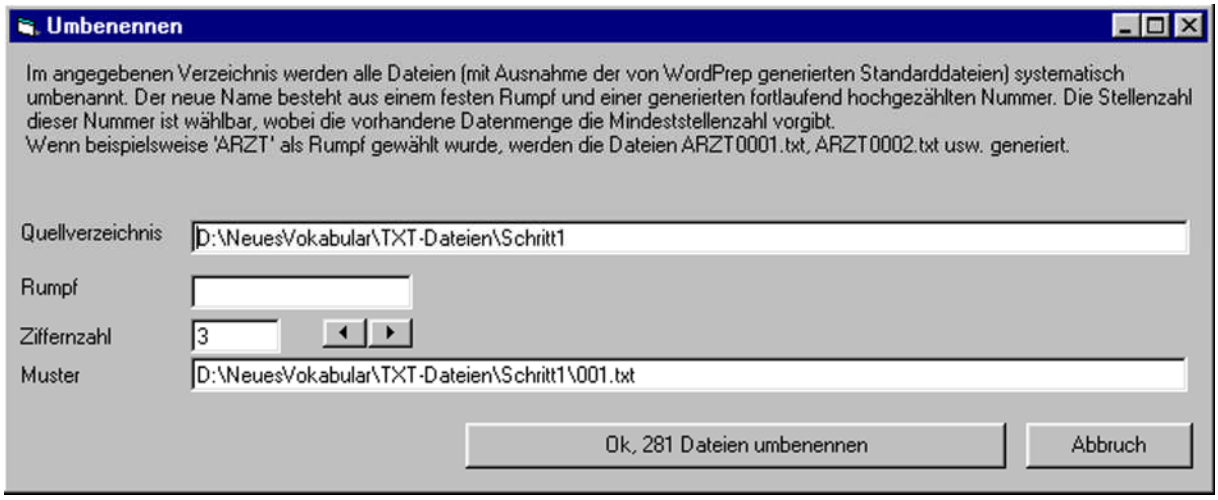
Diese Liste ist dann noch bezüglich Inhalt zu bearbeiten. Ggf. ist die gesprochene Form mit Tool **MODFYLST** zu ergänzen.

C) Sonstige Tools

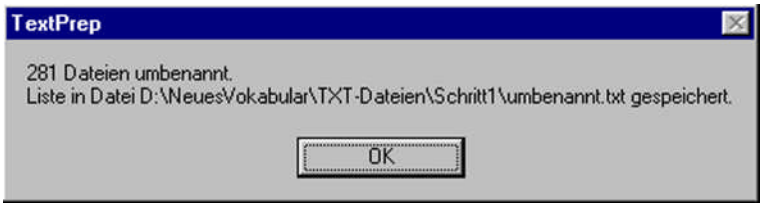
15. RENAME

Manche Quelldateien tragen in ihrem Dateinamen Hinweise auf Personen. Zur Anonymisierung ist es daher wünschenswert, diese Dateien umzubenennen. Das Tool RENAME ändert alle Dateinamen des Quellverzeichnisses in einen vorgegebenen Namen, ergänzt mit einer fortlaufenden Zahl.

Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool RENAME aus und klickt auf **Ok.** Es erscheint folgendes Menu:



Der Gesamtdateiname (Rumpf + fortlaufende Nummer) darf 8 Zeichen nicht überschreiten. Das Tool errechnet aus der Gesamtzahl von Dateien die Anzahl Stellen, die die fortlaufende Nummer benötigt. (In unseren Beispiel 3 Stellen).
Der noch einzugebende Rumpf (z.B. „Datei“) darf daher 1 bis max. 5 Zeichen betragen. Nach Eingabe dieses Rumpfes klickt man auf **Ok**, das Tool nimmt die Umbenennung vor und meldet sich am Ende sinngemäß:



Eventuell unterschiedliche Dateierweiterungen (z.B. .txt, .doc, .rtf usw.) bleiben unverändert!

16. COMBFILE

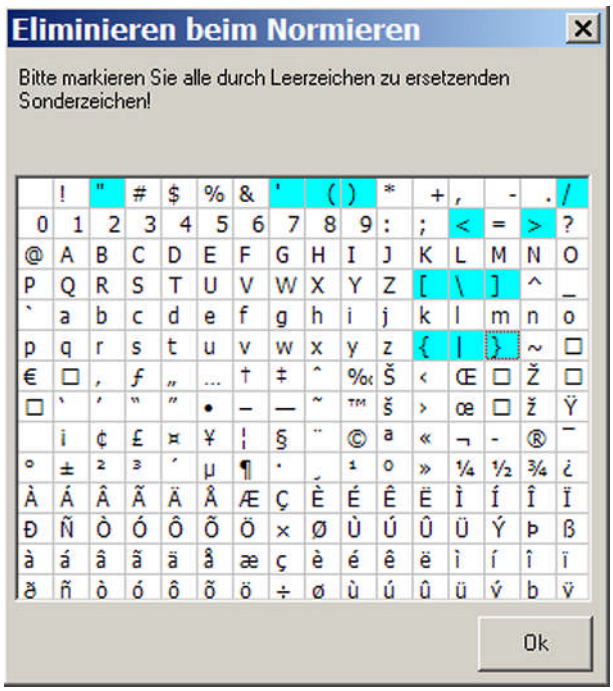
Arbeitet man anfänglich mit vielen kleinen Dateien, um z.B. mit CONTEXT möglichst viele Worttreffer in ihrem Kontext sehen zu können, wird es doch von manchen gewünscht, wenn man mit dem Nuance-Tool VOCTOOL eine Wortliste der unbekannten Worte erstellt, aus allen Quelldateien 1 große Datei oder wenige Dateien zu bilden.

Das vorliegende Tool COMBFILE fasst alle Dateien des Quellverzeichnisses zu einer beliebigen Anzahl von Dateien zusammen und speichert diese im Zielverzeichnis unter dem/den Namen „Rumpfname-xx.txt“ ab, wobei xx eine fortlaufende Zahl ist.

Es sollte sinnvollerweise vor der Ausführung dieses Tools darauf geachtet werden, dass alle Dateien des Quellverzeichnisses vom Typ TXT sind!

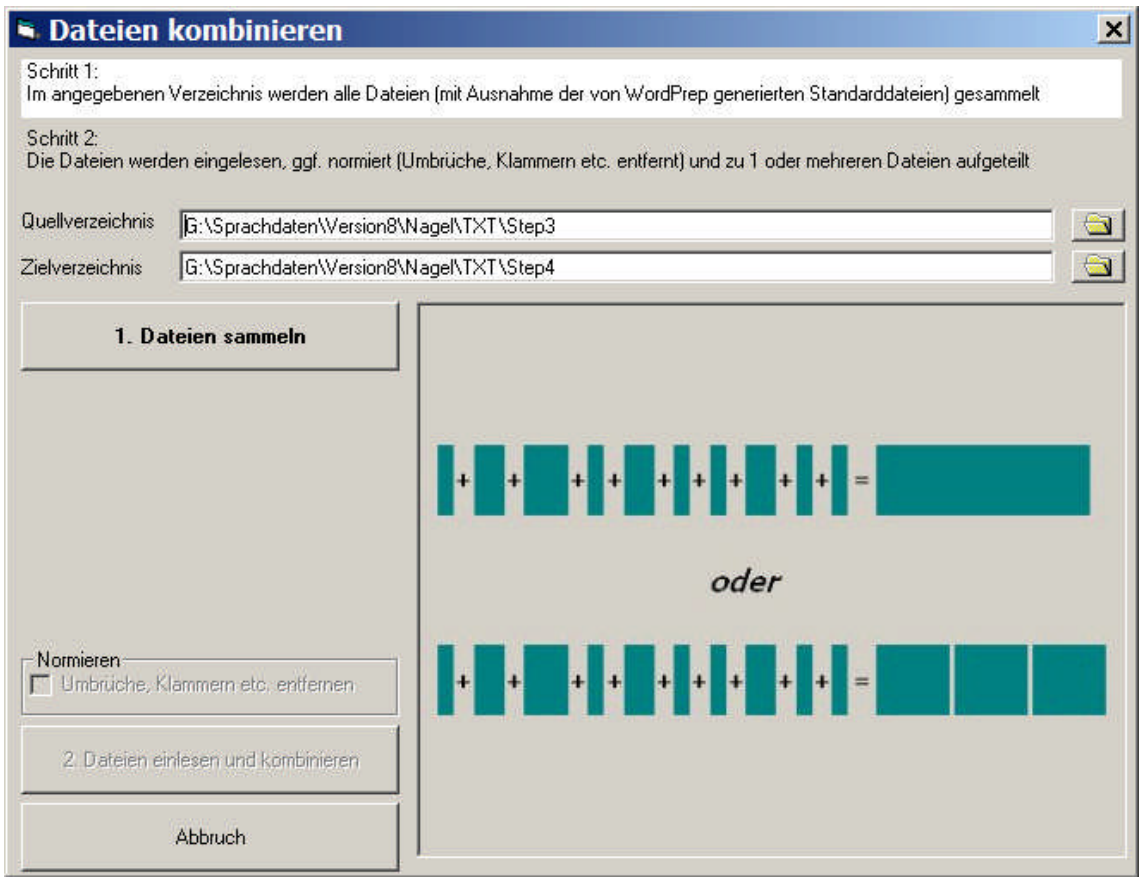
Bevor man das Tool COMBFILE startet, muss (zumindest beim ersten Aufruf des Tools die zugehörige Tabelle **Eliminieren beim Normieren** bearbeitet werden.

Man hat nämlich bei der Ausführung von COMBFILE die Option, dass die zu erstellende Datei(en) auch „normiert“ wird/werden. „Normieren heißt, dass alle überflüssigen Zeichen wie z.B., Klammern, Hochkommas, usw. werden entfernt, ebenso überflüssige Leerzeichen. Der Abstand der einzelnen Wörter wird auf 1 Leerzeichen reduziert. Die Tabelle „Eliminieren beim Normieren“ kennzeichnet alle Zeichen, die eliminiert bzw. durch ein Leerzeichen ersetzt werden.

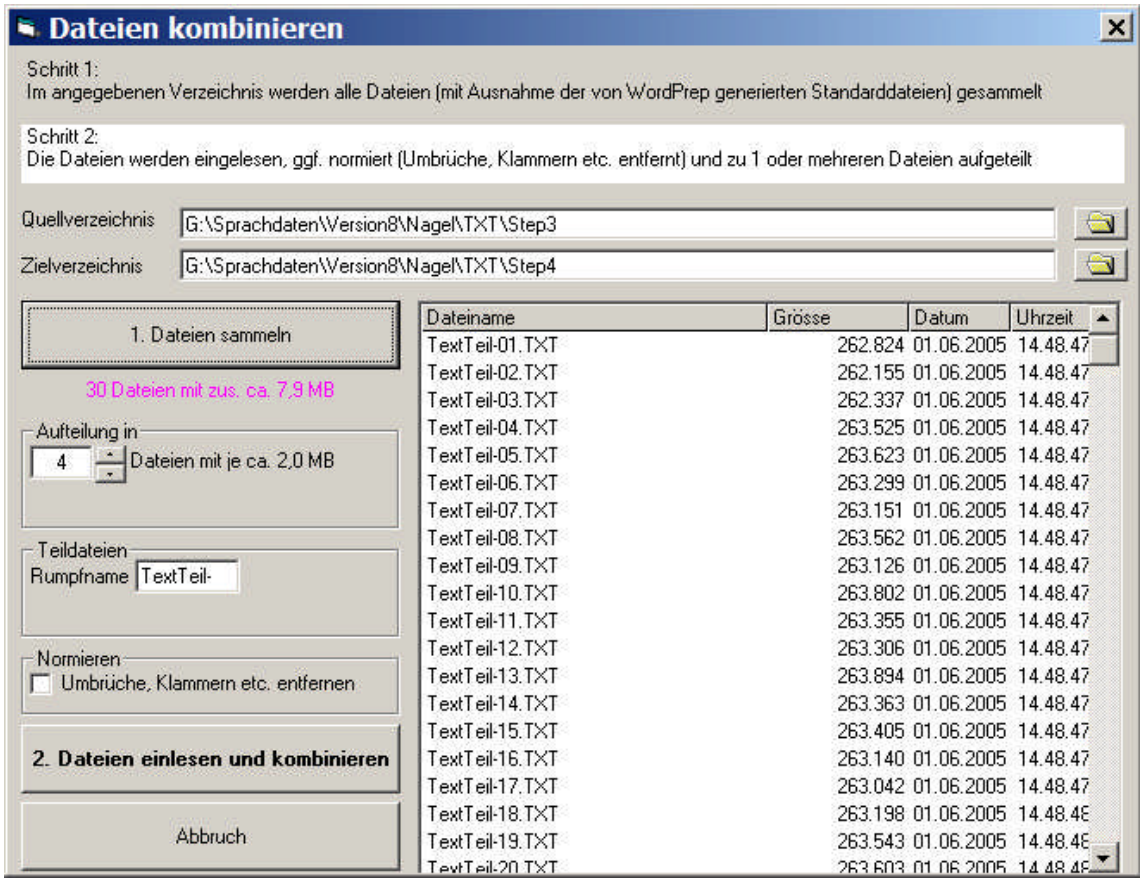


Achtung: Markieren Sie in dieser Tabelle keine Satzzeichen! Sonst werden diese durch ein Leerzeichen ersetzt und es gehen für die Bigrammstatistiken des Dragon NaturalVocTools wertvolle Informationen verloren.

Im ersten Schritt gibt man das Quell- und Zielverzeichnis an. Und klickt dann auf die Taste **1. Datei sammeln.**



Im nachfolgenden Schritt gibt man den Anfang des gewünschten Dateinamens ein, den sog. Rumpfnamen, spezifiziert, ob man eine Normierung des Textes wünscht (Entfernen von Klammern und Umbrüchen), und klickt dann auf die Taste **2. Dateien einlesen und kombinieren**.



Zum Start des Tools wählt man auf dem Hauptmenu von TextPrep das Tool COMBFILE aus und klickt auf **Ok.**

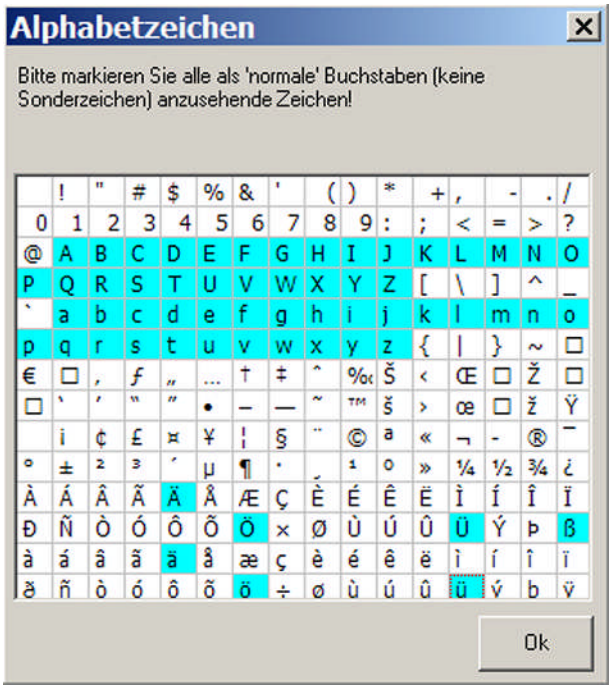
Die erstellte(n) Datei(en) wird/werden unter dem Namen „Rumpfname-xx.txt“ im Zielverzeichnis gespeichert.

17. REMOVERR

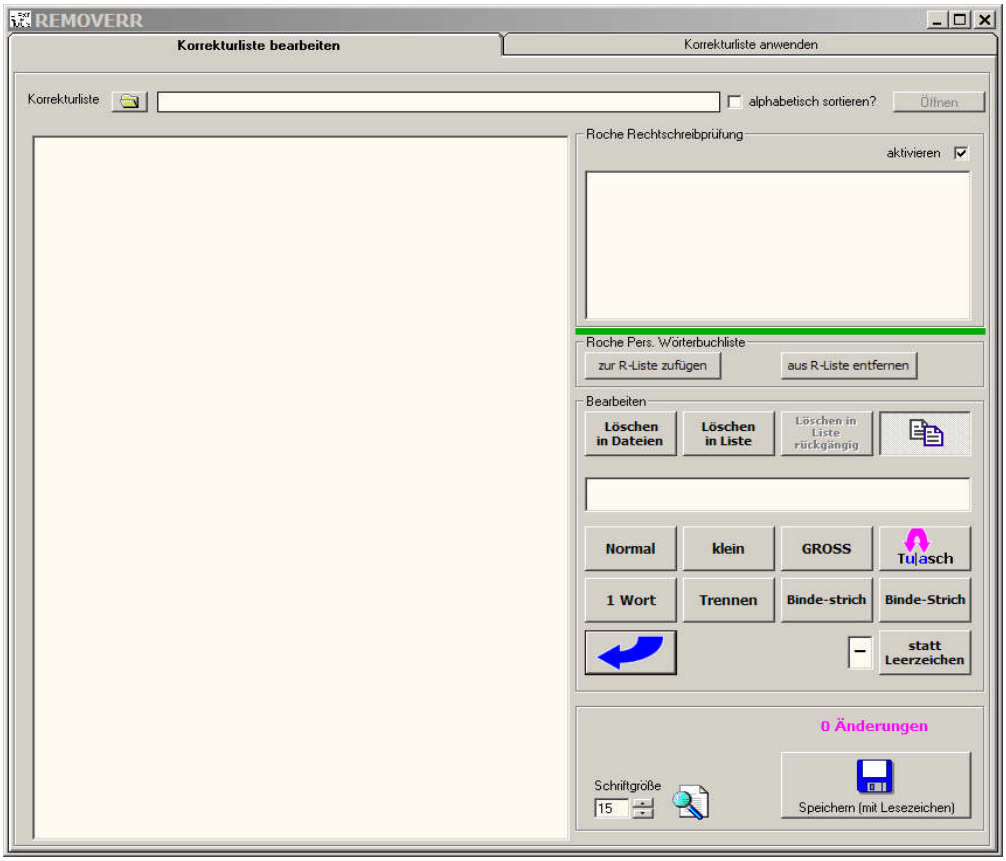
Um den Kontext neuer Begriffe nutzen zu können, ist es beim Erstellen neuer Fachvokabulare bisher ein zeitaufwendiger und beschwerlicher Schritt gewesen, anhand der Wortliste des Nuance VocTools mittels z.B. der MS Word Ersetzen-Funktion alle Schreibfehler im Quelltext zu korrigieren.

Das Tool REMOVERR hilft, die fehlerhaften Worte in jeder Wortliste auf einfachste Weise zu korrigieren. **REMOVERR überträgt dann am Ende alle diese Änderungen automatisch in den Quelltext!**

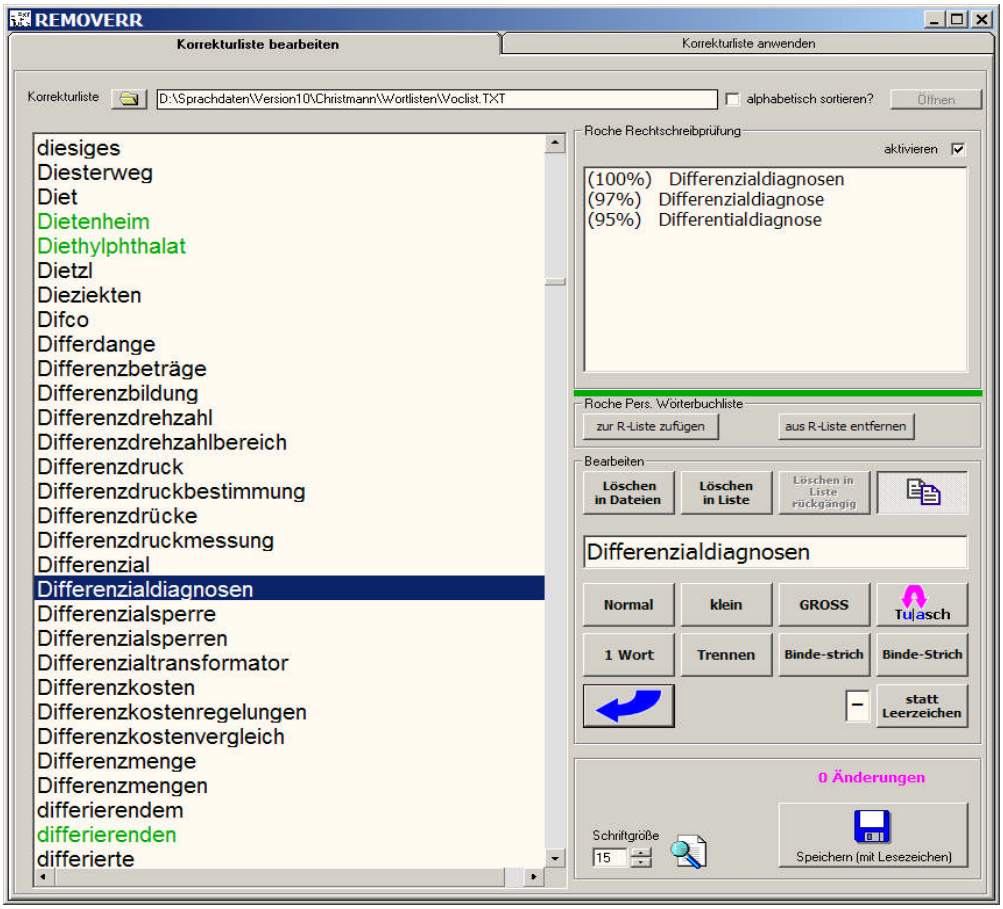
Beim allerersten Aufruf dieses Tools klickt man, nachdem man im TextPrep-Hauptmenu das Tool REMOVERR aktiviert hat, auf „Zugehörige Tabelle bearbeiten“, um festzulegen, welche Zeichen zum Alphabet gehören sollen.



Danach ruft man das Tool REMOVERR durch Klicken auf OK auf. Es erscheint folgendes, am Anfang noch leeres Fenster:



Im Feld „Korrekturliste“ wählt man die TXT-Wortliste aus, bei der man fehlerhafte Worte korrigieren möchte.



Man geht nun Wort für Wort durch diese Liste. Alle Wörter, die auch in der Referenzdatei der Roche Rechtschreibprüfung enthalten sind, sind **grün** hinterlegt. Sie müssen also in aller Regel nicht auf Rechtschreibfehler überprüft werden! Ist ein Begriff dabei, der geändert oder gelöscht werden soll, klickt man auf diesen Begriff, der daraufhin im Bearbeitungsfenster rechts wiederholt wird. Man hat nun folgende, einfache Bearbeitungsoptionen:

Normal
Der erste Buchstabe des Wortes wird groß geschrieben, der Rest klein.

klein
Alle Buchstaben des Wortes werden klein geschrieben.

GROSS
Alle Buchstaben des Wortes werden groß geschrieben (z.B. ADAC).

Tu|asch
Die beiden Buchstaben vor und nach dem Cursor werden vertauscht.

1 Wort
Der Begriff, meist ein Begriff, der einen Bindestrich enthält, wird als 1 Wort geschrieben (z.B. Abfahrts-zeit=>Abfahrtszeit).

Trennen
Bei einem Bindestrichbegriff wird der Bindestrich (genau genommen: alle Zeichen, die nicht zum Alphabet gehören) durch ein Leerzeichen ersetzt, der Begriff also in 2 Worte getrennt (z.B. manisch-gestört=>manisch gestört).

Binde-strich
In das Wort wird an der Stelle, an der der Cursor steht, ein Bindestrich eingefügt. Der Wortteil nach dem Bindestrich ist klein geschrieben.

Binde-Strich
In das Wort wird an der Stelle, an der der Cursor steht, ein Bindestrich eingefügt. Der Wortteil nach dem Bindestrich ist groß geschrieben.

Löschen in Dateien
Der Begriff wird in den Quelltexten ersatzlos gelöscht.

Löschen in Liste
Der Begriff wird (nur) in der Wortliste gelöscht

Löschen in Liste rückgängig
Mit dieser Taste wird der letzte Löschen-in-Liste Befehl rückgängig gemacht.

Die Taste (einmal am Anfang EIN oder AUS) stellt jeden Begriff, der im linken Fenster angeklickt wird, automatisch in die Windows Zwischenablage. Wenn man nun parallel das TextPrep-Tool CONTEXT geöffnet hat, sieht man unmittelbar die Textumgebungen, in denen dieser Begriff vorkommt. Hierbei ist es sehr übersichtlich, wenn das Tool CONTEXT auf einem zweiten Monitor läuft.

Blauer Pfeil 

Beim Klicken auf den blauen Pfeil wird das geänderte Wort in Form einer Ersetzungsanweisung (ähnlich wie beim Tool REMOVE) in die Wortliste übernommen. Ein Klick auf das nächste Wort bewirkt das Gleiche wie ein Klick auf den Blauen Pfeil.

Besteht ein Begriff aus mehreren Wörtern und möchte man diese mit einem Bindestrich (oder mit einem anderen Zeichen) verbinden, geschieht das durch einen Klick auf diese Taste.

Grüner Strich
Mit dem grünen Strich lässt sich die Größe des Fensters der Roche Rechtschreibprüfung stufenlos verändern.

Roche Rechschreibprüfung

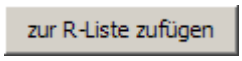
In diesem Fenster werden alternative Schreibweisen für das ausgewählte Wort angezeigt. Ist die richtige bzw. gewünschte Schreibweise dabei, kann man diese durch Doppelklick übernehmen.

Sonstige Optionen

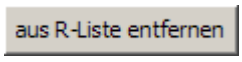
- Man kann manuell das Wort beliebig ergänzen oder verändern. D.h., man erzeugt eine Ersetzungsanweisung.
- Es kann auch das Ausgangswort, so wie es in der Wortliste erscheint, durch einen Doppelklick verändert werden. Dies ist dann sinnvoll, wenn z.B. der Begriff eine Abkürzung ist, der Punkt jedoch nach der Abkürzung fehlt. Dann kann man z.B. aus „zusätzl“ zuerst „zusätzl.“ machen und dann im zweiten Schritt die Ersetzungsanweisung „zusätzl.=>zusätzlich“ erzeugen. Dadurch werden dann im Quelltext alle Vorkommen von „zusätzl.“ durch „zusätzlich“ ersetzt.
- Es können mehrere Begriffe zusammenhängend markiert und gelöscht werden.
- Da Wortlisten sehr umfangreich sein können, kann man den momentanen Verarbeitungsstand speichern **„Speichen (mit Lesezeichen)“**. Die Wortliste wird dann mit der Dateierweiterung KOR abgespeichert! Beim erneuten Aufruf dieses Tools wählt man die KOR-Datei aus und kann an der Stelle weitermachen, an der man die Verarbeitung unterbrochen hat.

- Zur Schonung der Augen kann man die Schriftgröße einstellen. Auch ist das Fenster in der Größe anpassbar. Man muss nur den unteren Rand nach unten ziehen, um mehr Worte auf dem Bildschirm anzuzeigen!

Man kann während der Listenbearbeitung Begriffe dieser Liste "on the fly" in die Referenzdatei der Roche-Rechtschreibprüfung kopieren, um diese schrittweise anzureichern. Dazu markiert man 1 oder mehrere Begriffe der Wortliste und klickt dann auf das Feld



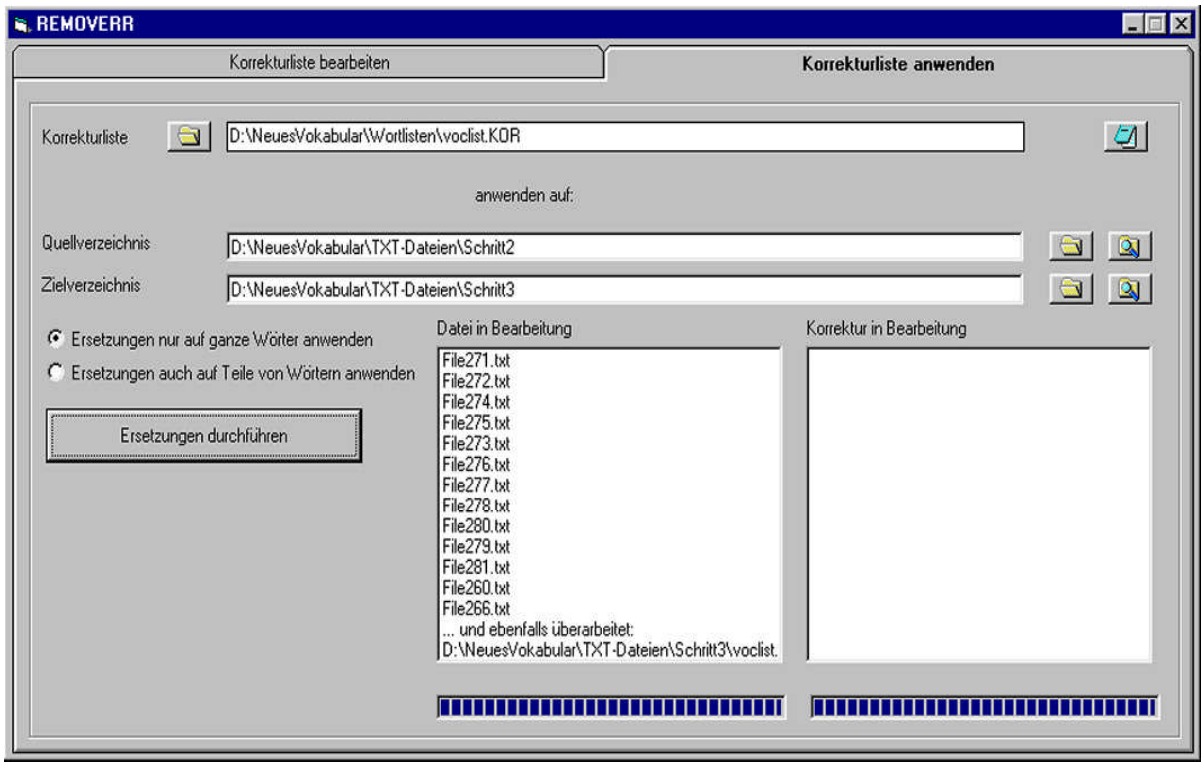
Ebenso kann man einen Begriff, der in dem Fenster Roche-Rechtschreibprüfung angezeigt wird, wieder löschen. Dazu den Begriff markieren und dann auf die Taste



klicken.

Wählt man am Ende der Bearbeitung, oder auch zwischendurch, die Taste **Speichern**, erscheint eine Tabelle mit allen Begriffen, die man der Roche-Referenzdatei hinzufügen bzw. löschen möchte. Diese Tabelle kann noch editiert werden.

Ist man mit der Bearbeitung der Wortliste fertig, wählt man das Fenster **Korrekturliste anwenden**.



Bei Klick auf **Ersetzungen durchführen** führt nun das Tool REMOVERR

a) die Änderungen in der bearbeiteten Wortliste durch, und
b) führt auch alle Änderungen im Quelltext durch.

Es sollte üblicherweise die Option **Ersetzungen nur auf ganze Wörter anwenden** ausgewählt sein.

Besonderheit

Dem Toolset TextPrep ist die Datei **AltZuNeu.txt** beigefügt. Diese setzt weitgehend die alte deutsche Rechtschreibung in die neue um. Mit Hilfe von REMOVERR und dieser Datei können also Quelltexte mühelos umgestellt werden!! Da in der Datei AltZuNeu.txt auch Silbenumsetzungen vorgenommen werden, sollte hier bei REMOVERR die Option **Ersetzungen auch auf Teile von Wörtern anwenden** ausgewählt sein.
Die Datei AltZuNeu.txt kann beliebig durch den Benutzer ergänzt/verändert werden.

Die mit REMOVERR bearbeiteten Wortlisten (KOR-Dateien) enthalten neben unveränderten Worten auch eine ganze Reihe von Ersetzungsanweisungen. Diese Ersetzungsanweisungen kann man nicht nur auf den Quelltext anwenden, sondern auch aufbewahren, um sie bei einem anderen Quelltext des gleichen Fachgebiets erneut anzuwenden. Dabei möchte man aber die unveränderten Worte aus den KOR-Listen entfernen, d.h. nur die Ersetzungsanweisungen behalten. Dies erreicht man mit dem Tool DELLINES mit der Option „Alle Zeilen löschen, in denen ein bestimmtes Textmuster NICHT vorkommt“ und dem Textstring „=>“ (ohne Anführungszeichen).

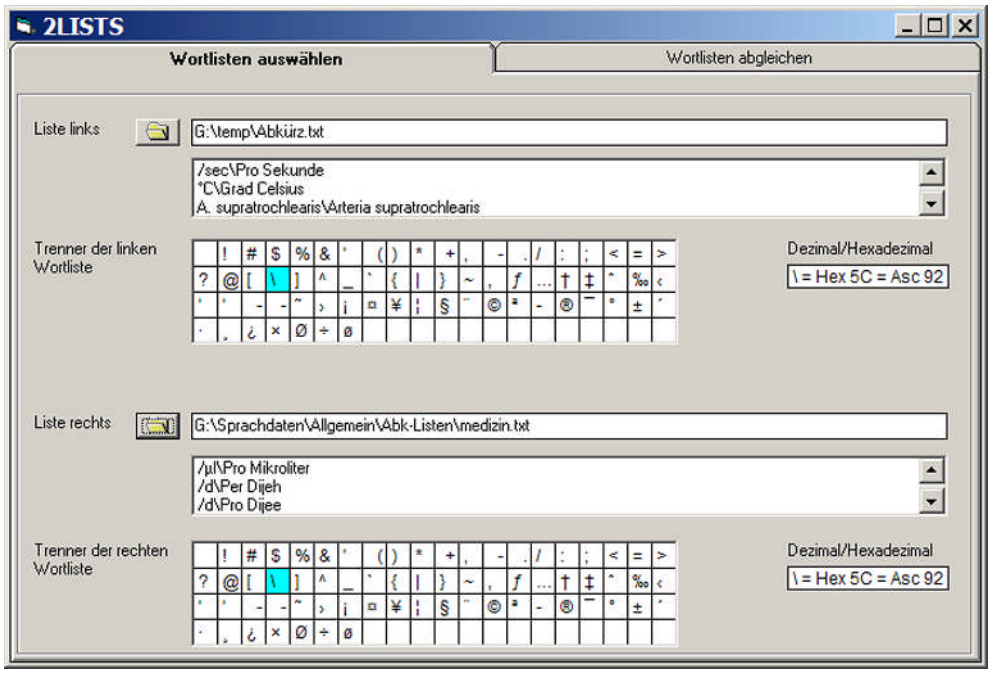
18. 2LISTS

Bei den erzeugten Wortlisten **Bindestrichbegriffe** („Mutter-Kind-Beziehung“) und den **Begriffen mit einem sog. Ergänzungsstrich** („Auf- und Umbau“) wird die gesprochene Form auf Wunsch automatisch generiert.

Bei Wortlisten wie z.B. **Abkürzungen** („Tabl.“)**, Maßeinheiten (**„mg/l“**),** und bei sog. **Großbuchstabenworte** („ADAC“) ist die gesprochene Form unbedingt erforderlich, kann jedoch nicht automatisch erzeugt werden. Da diese Listen jedoch sehr lang sein können, ist es sehr zeitraubend, diesen Listen die gesprochene Form hinzuzufügen.

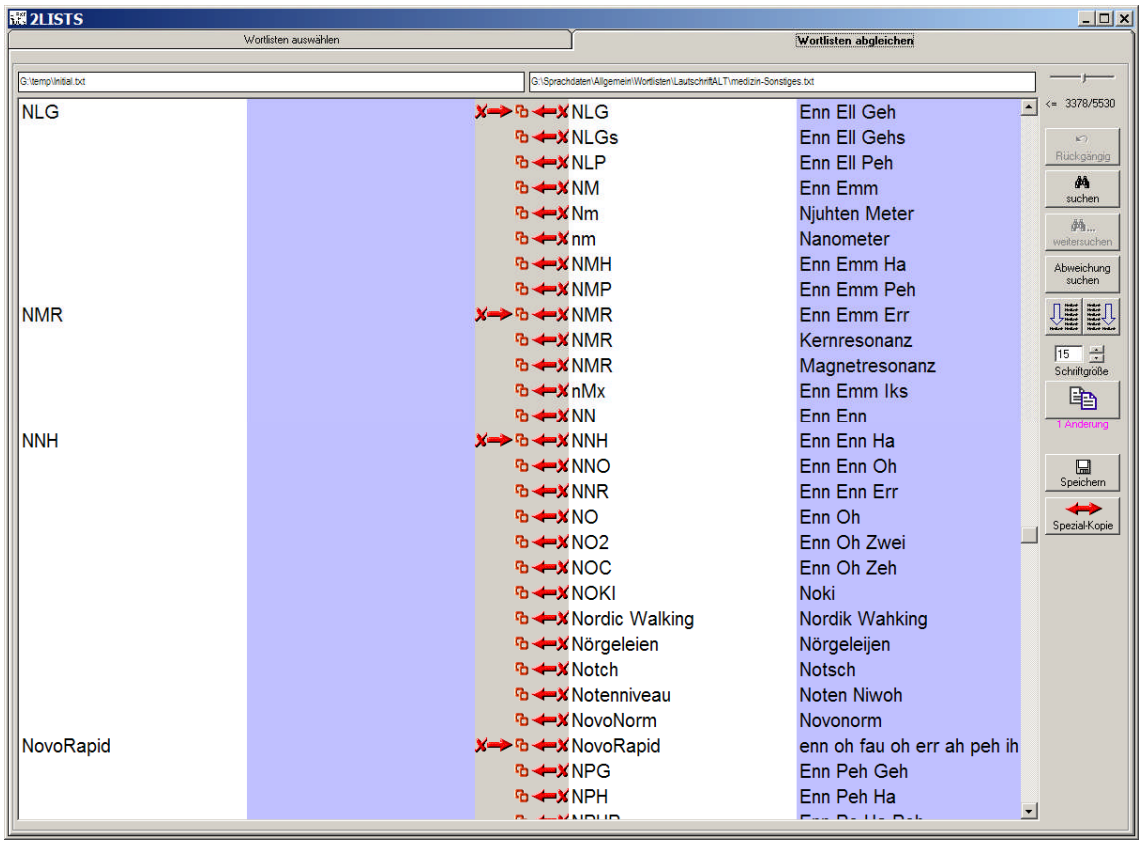
Es kommt noch hinzu, dass beim Erstellen von Fachvokabularen mehr als 80 Prozent dieser Begriffe und ihre gesprochene Form immer wieder vorkommen. Das Tool 2LISTS erlaubt nun, bei einer neu erstellten Liste die gesprochene Form aus früheren Listen auf leichte und sehr zeitsparende Weise zu übernehmen.

Zum Aufruf des Tools wählt man das Tool im Hauptmenu aus und bestätigt mit dem Auswahlknopf „OK“. Es erscheint das Auswahlmenu **Wortlisten auswählen“:**



Man gibt nun im Feld **Liste links** die Datei an, der man gesprochene Formen hinzufügen möchte. Im Feld **Liste rechts** spezifiziert man die Datei, die (aus früheren Fachvokabularerstellungen) möglichst viele Begriffe mit ihrer gesprochenen Form enthält. Als **Trenner** zwischen der geschriebenen und der gesprochenen Form gibt man bei Dragon NaturallySpeaking den **Backslash** an. Danach klickt man auf das Karteiblatt **Wortlisten abgleichen**.

Daraufhin wird der Anfang beider Wortlisten abgebildet. Hier ein Ausschnitt aus 2 Listen, wie es beispielhaft aussehen könnte.



Links die „neue“ Liste ohne gesprochene Formen und rechts die Liste mit den gesprochenen Formen. Auf beiden Bildschirmhälften kommt zuerst ein Bereich für die geschriebene Form (weißer Hintergrund) gefolgt vom Bereich für die gesprochene Form (getönter Hintergrund).

Die roten Funktionszeichen zwischen den beiden Dateien haben folgende Bedeutung:

Lösche in der linken bzw. der rechten Datei den Inhalt der Zeile (geschriebene und gesprochene Form)

Kopiere den Inhalt dieser Zeile der linken Datei (geschriebene und gesprochene Form) in die rechte Datei

Dupliziere in beiden Dateien diese Zeile (geschriebene und gesprochene Form)

Kopiere den Inhalt dieser Zeile der rechten Datei (geschriebene und gesprochene Form) in die linke Datei

Bedeutung der Tastenfelder

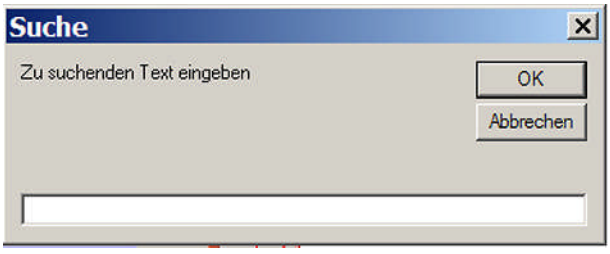
Am rechten senkrechten Bildschirmrand sind folgende Tastenfelder:

Dieser Schieber regelt das Breitenverhältnis der Felder für die geschriebene und gesprochene Form. Durch Verschieben des Reglers ändert sich das Breitenverhältnis dieser Felder.

Aus diesen Angaben entnimmt man die Zeilennummer (1. Zeile des aktuellen Bildschirmauschnitts und die Gesamtzahl aller Zeilen. Im Beispiel ist die erste auf dem Bildschirm sichtbare Zeile die 1396. Zeile von insgesamt 2289 Zeilen.

Mit dieser Taste kann die letzte Aktion (Veränderung) rückgängig gemacht werden.

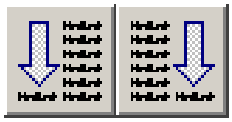
Man kann hiermit über einen Suchvorgang zu einer bestimmten Textstelle springen (sofern vorhanden). Bei der Auswahl dieser Taste öffnet sich ein Fenster zur Eingabe des Suchbegriffes:



Nach Eingabe des Suchbegriffes klickt man auf **OK** oder **Abbrechen**. Daraufhin wird die nächste Stelle angezeigt, die den Suchbegriff enthält.

Damit wird der nächste Treffer in der Liste angezeigt, der zuvor bei **Suchen** eingegeben wurde.

Manches Listen haben im Vergleich nur geringfügige Abweichungen, fast alle Listeneinträge sind identisch. Mit dieser Taste springt man zur nächsten Zeile, bei der sich die beiden Listen unterscheiden.



Mit diesen Tasten springt man zum nächsten Listenelement in der linken bzw. rechten Wortliste.

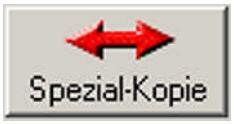
Zur besseren Lesbarkeit der Begriffe, oder um mehr Zeilen auf dem aktuellen Bildschirm darstellen zu können, kann hiermit die Schriftgröße verändert werden.

Diese Taste (einmal am Anfang EIN oder AUS) stellt jeden Begriff, der im linken Fenster angeklickt wird, automatisch in die Windows Zwischenablage. Wenn man nun parallel das TextPrep-Tool CONTEXT geöffnet hat, sieht man unmittelbar die Textumgebungen, in denen dieser Begriff vorkommt. Hierbei ist es sehr übersichtlich, wenn das Tool CONTEXT auf einem zweiten Monitor läuft.

Mit dieser Taste werden **beide** Dateien gespeichert (ohne „Schließen, d.h. die Bearbeitung kann an dieser Stelle fortgesetzt werden).

Hat man anfangs nur 1 Datei geöffnet, was möglich ist, und hat man in die 2., anfangs leere Datei etwas kopiert, wird man aufgefordert, einen Ort und Namen dieser 2. Datei anzugeben, um

sie zu speicher. Die erste Datei wird ohne Aufforderung ebenfalls gespeichert.



Diese Taste ruft ein Aktionsfenster zum Massenkopieren auf, da Wortlisten sehr lang sein können und ein einzelnes Kopieren von Zeilen zu zeitaufwendig sein kann.

Alle von Links nach Rechts ggf. mit Überschreiben vorhandener Einträge

Alle Zeilen der linken Datei incl. eventueller gesprochener Formen werden auch in die rechte Datei eingefügt. Sollten dort z.T. diese Begriffe schon vorkommen (ggf. mit leicht abweichender gesprochener Form), werden diese überschrieben

Alle von Links nach Rechts ohne Überschreiben

Alle Zeilen der linken Datei incl. eventueller gesprochener Formen werden auch in die rechte Datei eingefügt. Sollten dort z.T. diese Begriffe schon vorkommen, werden diese **nicht** überschrieben

Nur gesprochene Form von Links nach Rechts wenn geschriebene Form rechts vorhanden und gesprochene Form leer

Mit dieser Taste kann man alle gesprochenen Formen von links nach rechts übertragen, wenn in der rechten Datei die geschriebene Form schon existiert, jedoch noch keine gesprochene.

Linke Seite löschen, wenn linke und rechte geschriebene Form identisch sind

Damit erzeugt man auf der linken Seite eine Wortlistendatei, die in keiner Zeile eine Übereinstimmung mit der anderen Datei hat. Dies kann hilfreich sein, wenn man diese Datei in einem zweiten Schritt der anderen Datei hinzufügen möchte, ohne Duplikate zu erzeugen.

Alle von Rechts nach Links ggf. mit Überschreiben vorhandener Einträge

Alle Zeilen der rechten Datei incl. eventueller gesprochener Formen werden auch in die linke Datei eingefügt. Sollten dort z.T. diese Begriffe schon vorkommen (ggf. mit leicht abweichender gesprochener Form), werden diese überschrieben

Alle von Rechts nach Links ohne Überschreiben

Alle Zeilen der rechten Datei incl. eventueller gesprochener Formen werden auch in die linke Datei eingefügt. Sollten dort z.T. diese Begriffe schon vorkommen, werden diese **nicht** überschrieben

Nur gesprochene Form von Rechts nach Links wenn geschriebene Form links vorhanden und gesprochene Form leer

Mit dieser Taste kann man alle gesprochenen Formen von rechts nach links übertragen, wenn in der rechten Datei die geschriebene Form schon existiert, jedoch noch keine gesprochene.

Rechte Seite löschen, wenn linke und rechte geschriebene Form identisch sind

Damit erzeugt man auf der rechten Seite eine Wortlistendatei, die in keiner Zeile eine Übereinstimmung mit der anderen Datei hat. Dies kann hilfreich sein, wenn man diese Datei in einem zweiten Schritt der anderen Datei hinzufügen möchte, ohne Duplikate zu erzeugen.

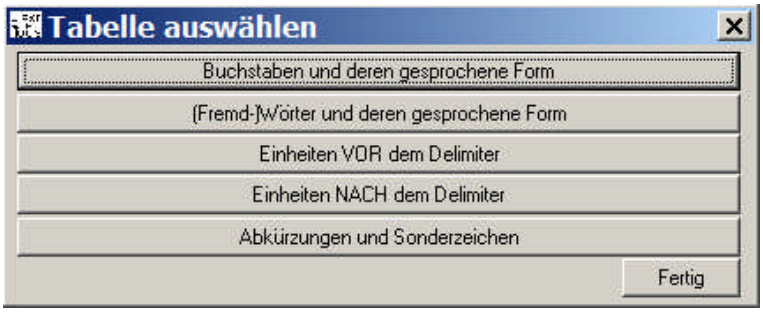
Achtung:

- 2LISTS sortiert die eingelesenen Wortlisten automatisch
- 2LISTS entfernt Leerstellen am Anfang und Ende jeder Zeile automatisch
- Man kann auch nur 1 Datei einlesen. Man kann dann ausgewählte Wörter auf die 'leere' Seite kopieren. Beim Speichern wird man immer auch nach dem Namen u. Speicherort der 2. Datei gefragt, auch wenn diese leer geblieben ist. Im letzteren Fall "Abbrechen" auswählen.

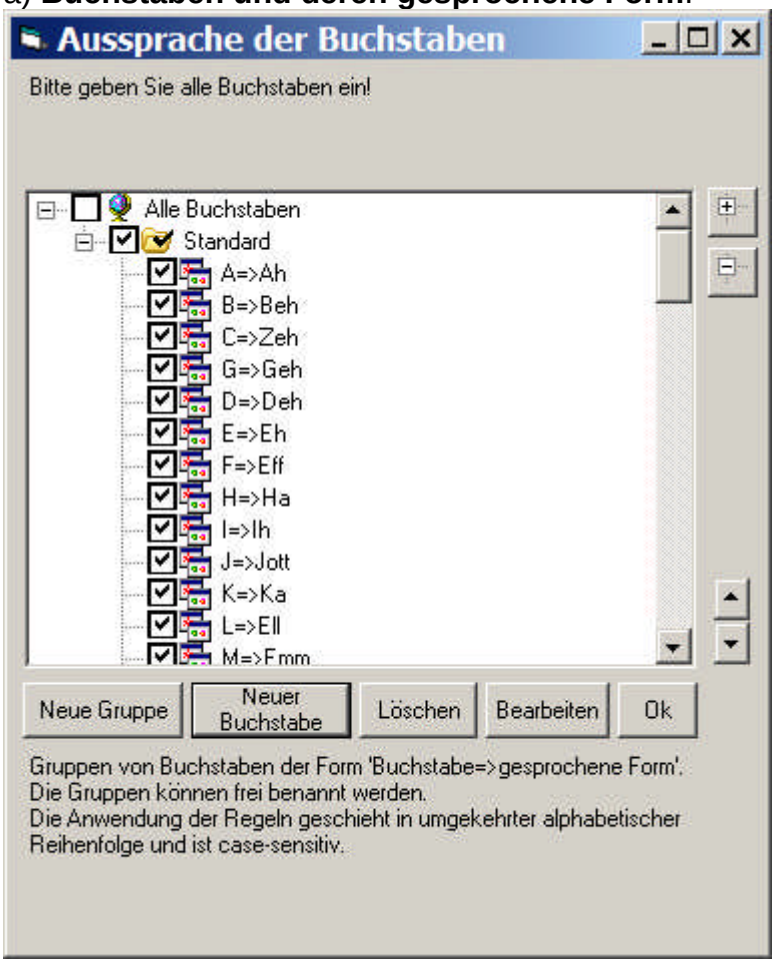
19. MODFYLST

Mittels dieses Tools kann man in einer Wortliste die gesprochene Form erzeugen und/oder selektiv die gesprochene oder die geschriebene Form bearbeiten.

Zunächst werden vor Aufruf dieses Tools folgende Tabellen ausgewählt und erstellt oder ggf. angepasst, damit auch in besonderen Fällen TextPrep die gesprochene Form weitestgehend automatisch erstellt:



a) **Buchstaben und deren gesprochene Form**:



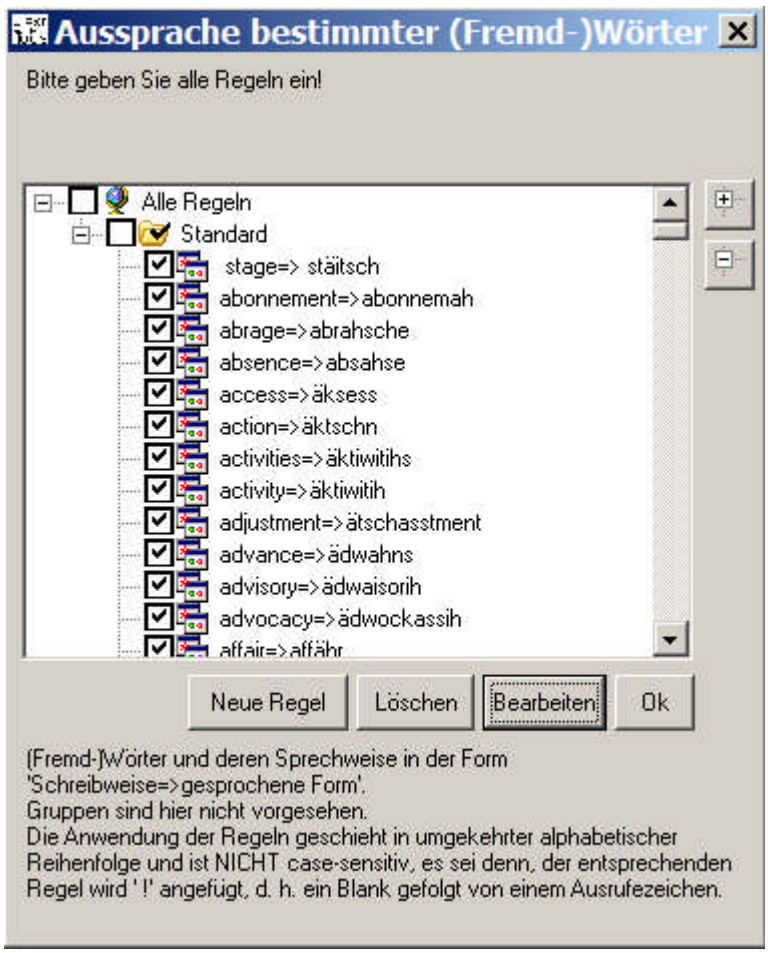
Hier gibt man für alle Buchstaben an, wie sie bei Initialwörtern diktiert werden, d.h. lautmalerisch geschrieben werden.

Format: **A=>Ah**
Beispiel: für das Initialwort "ADAC" wird dadurch die gesprochene Form "Ah Deh Ah Zeh" generiert.

In der mitgelieferten Datei "Einstellungen.txt" ist eine gesprochene Form vorgeschlagen. Diese muss überprüft werden. Ggf. z.B. ß=>Beta" in "ß=>"ess zett" o.ä. ändern.

b) **(Fremd-)Wörter und deren gesprochene Form**

Im Tool MODFYLST wird das Ziel verfolgt, die gesprochene Form weitgehend automatisch zu erzeugen. Dies soll auch für Fremdwörter und Wörter von nichtdeutschen Sprachen gelten. Die Tabelle "(Fremd-)Wörter und deren gesprochene Form" könnte z.B. so aussehen:



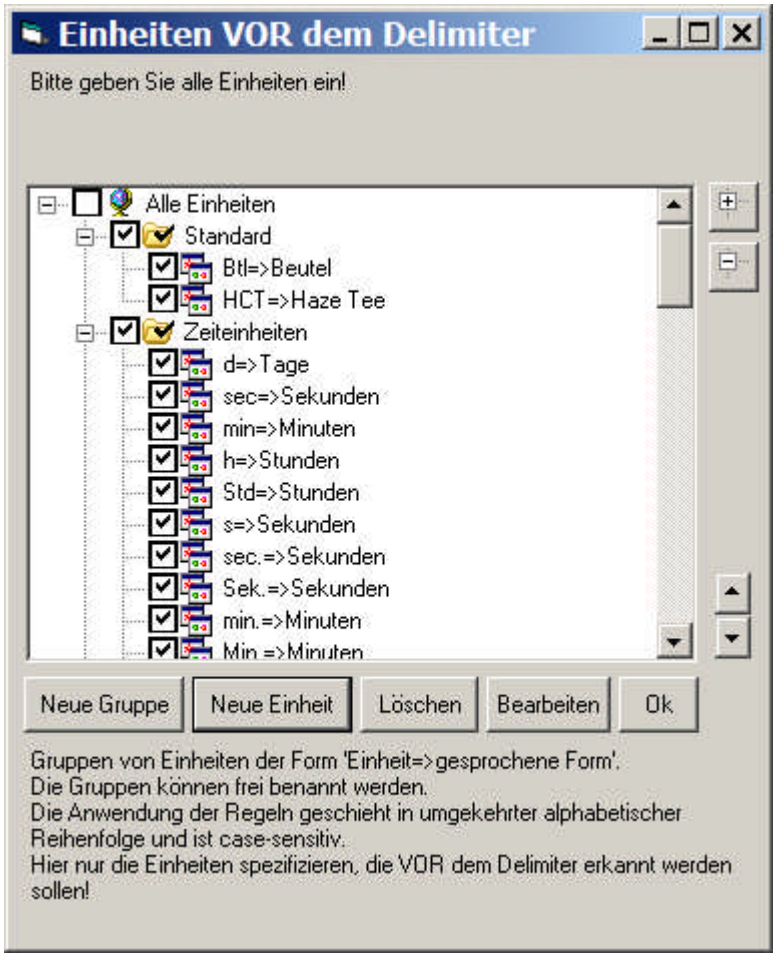
Es ist zu beachten, dass die in dieser Tabelle angegebenen Wörter Teil eines längeren Wortes sein können. Bei der Generierung der gesprochenen Form wird zuerst die geschriebene Form in das Feld "gesprochene Form" kopiert und dann der passende Teil gemäß dieser Tabelle ersetzt.

Auch Wörter mit Sonderzeichen, z.B. einem Akzent , können so eine gesprochene Form erhalten. Wenn man z.B. die Zeile "Lasègue=>lasähg" in die Tabelle eingibt, wird sie zwar als "Las[hex E8]gue=>lasähg" dargestellt, bei der Anwendung dieser Tabelle bekommt aber das Wort "Lasègue" die gesprochene Form "lasähg".

b) **Einheiten und deren gesprochene Form**
In diesen 2 Tabellen gibt man Maßeinheiten an, z.B. ml, km, und deren gesprochene Form an.

Format: **ml=>Milliliter**

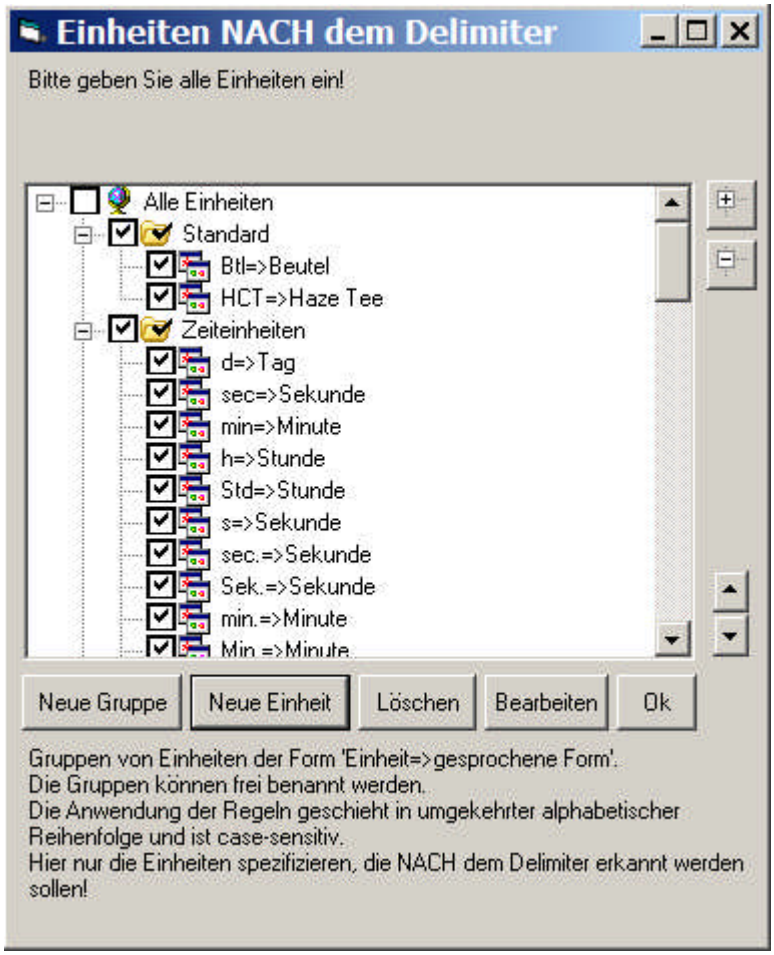
Enthält der Textstring einen Delimiter, üblicherweise in der deutschen Sprache der Schrägstrich ("/"), wird geprüft, ob vor oder nach dem Schrägstrich eine Maßeinheit steht. Wenn ja, wird die entsprechende gesprochene Form aus den Tabellen vor und nach dem Delimiter entnommen.



An Stelle des Schrägstrichs steht die gesprochene Form, wie sie bei der Auswahlseite der Datei angegeben wurde (s. weiter unten) .

Beispiel: aus der geschriebenen Form "km/h" wird die gesprochene Form "Kilometer Pro Stunde".

Es sind 2 Tabellen nötig, da vor dem Delimiter die Maßeinheiten üblicherweise im Plural angegeben werden, nach dem Delimiter im Singular.



d) **Abkürzungen und Sonderzeichen**

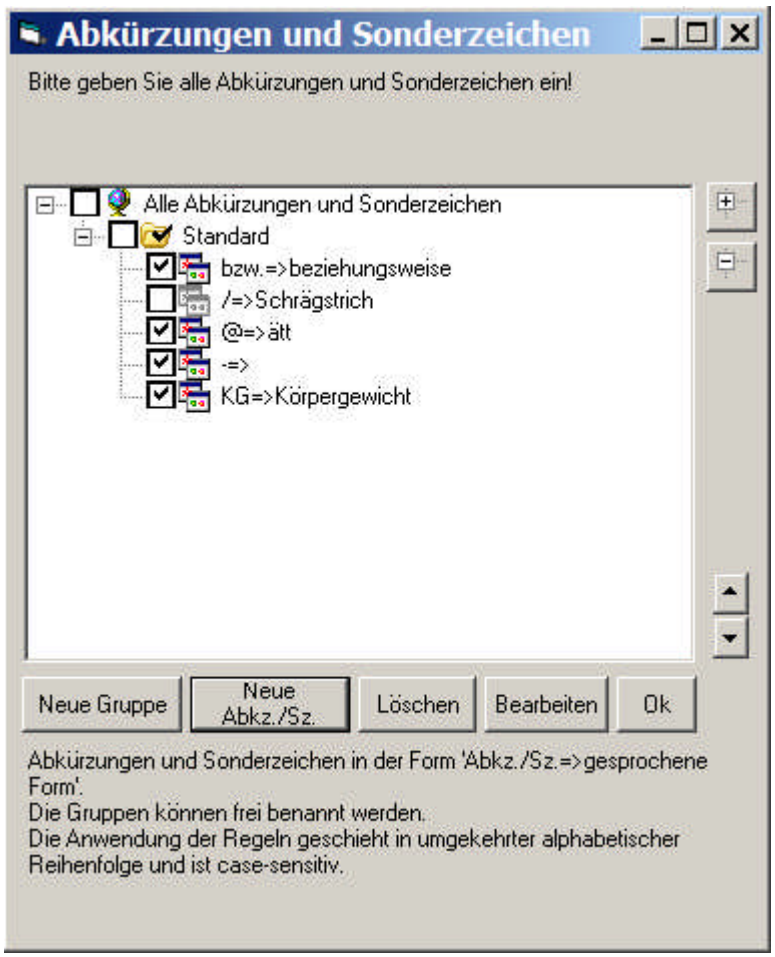
In dieser Tabelle werden alle Abkürzungen und Sonderzeichen aufgeführt, die anders gesprochen als geschrieben werden.

Beispiele:

bzw.=>beziehungsweise
@=>ätt
@=>Klammeraffe
KG=>Körpergewicht

Dies erlaubt, im Zusammenhang mit den anderen Tabellen z.B. folgende gesprochene Form zu generieren:

mg/kg KG/d --------> Milligramm Pro Kilogramm Körpergewicht Pro Tag



Sind die Tabellen nun erstellt bzw. angepasst, wird im nächsten Schritt im Feld "Liste" die zu bearbeitende Wortliste ausgewählt.

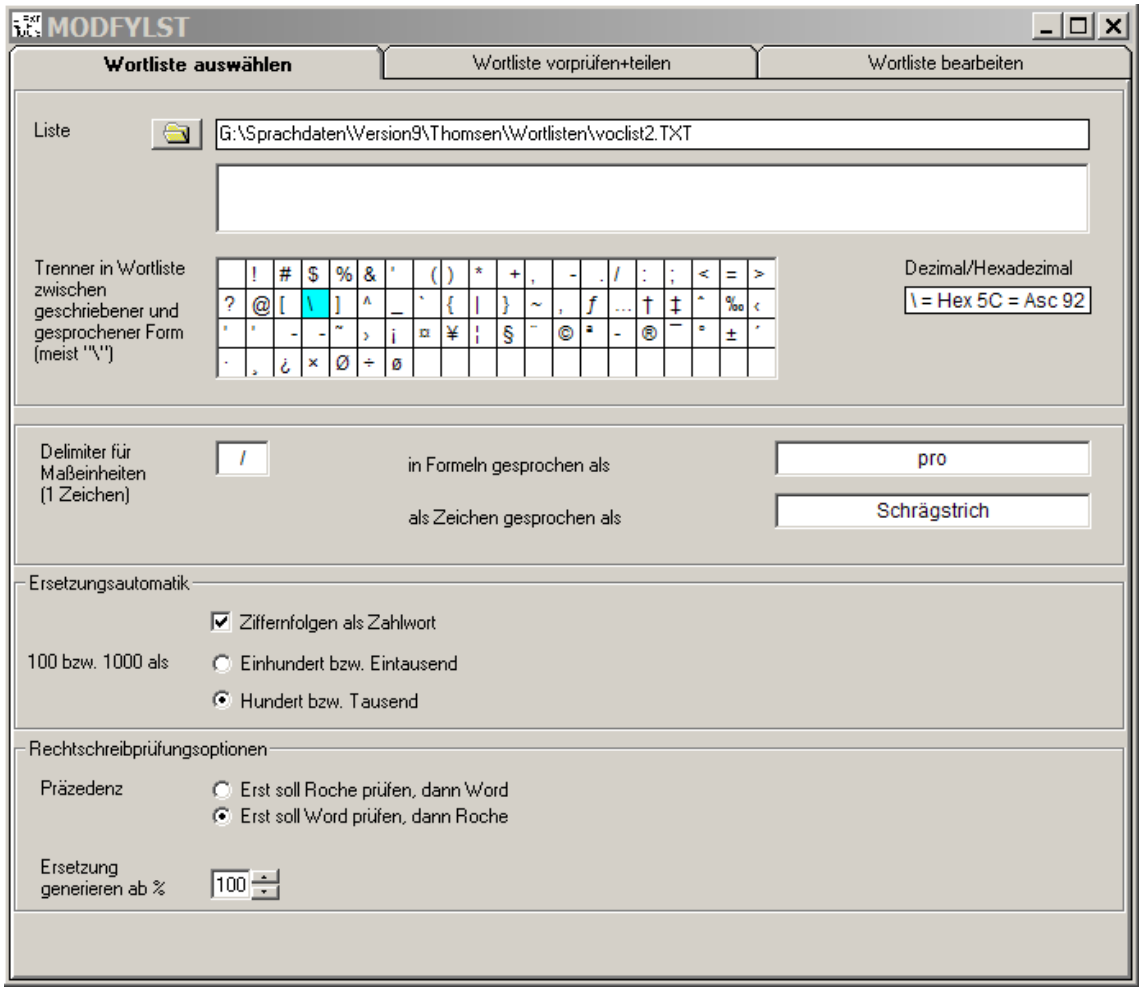
Als Trenner der Wortliste ist bei Dragon NaturallySpeaking der Backslash ("\") vorgeschrieben.

An dieser Stelle wird auch der Delimiter innerhalb der Maßeinheiten angegeben. Dies ist allgemein der Schrägstrich.
Innerhalb von Maßeinheiten wird im deutschen Sprachraum der Schrägstrich als "Pro" diktiert.

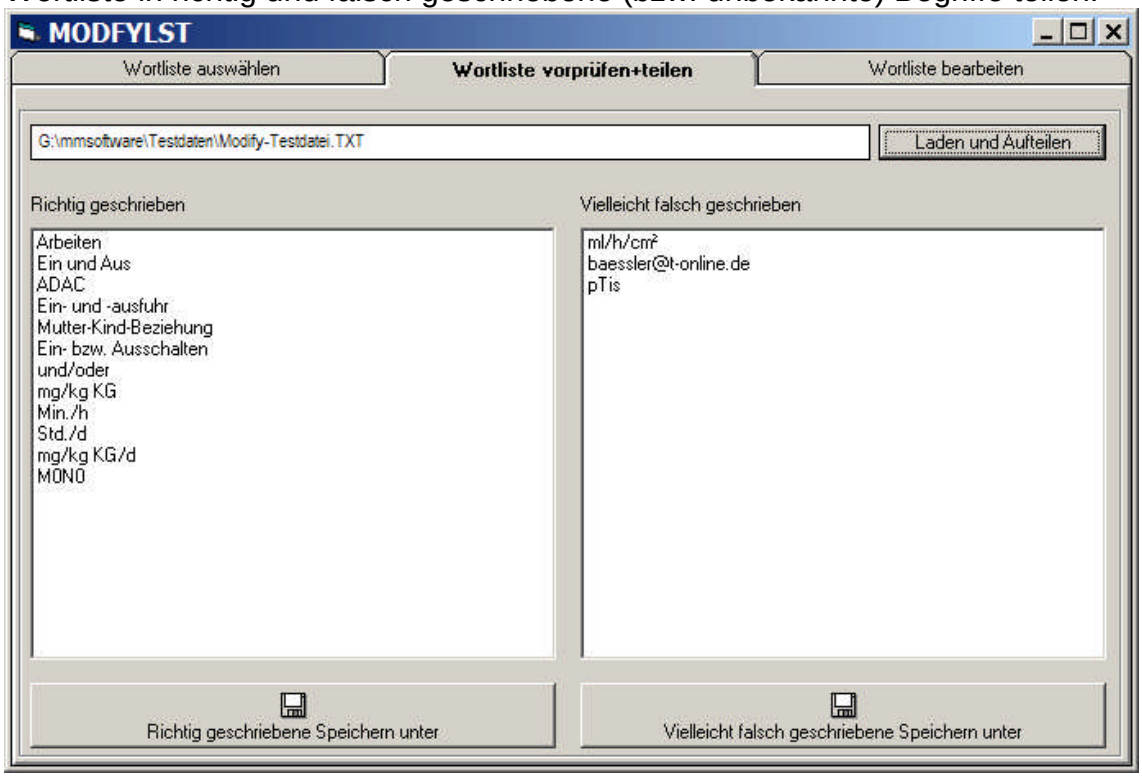
Beispiele: km/h = Kilometer pro Stunde, mg/cm² = Milligramm pro Quadratzentimeter.

Um TextPrep auch für andere Sprachen nutzen zu können, kann man für den Schrägstrich auch eine andere gesprochene Form angeben (z.B. "per" für englische Maßeinheiten).

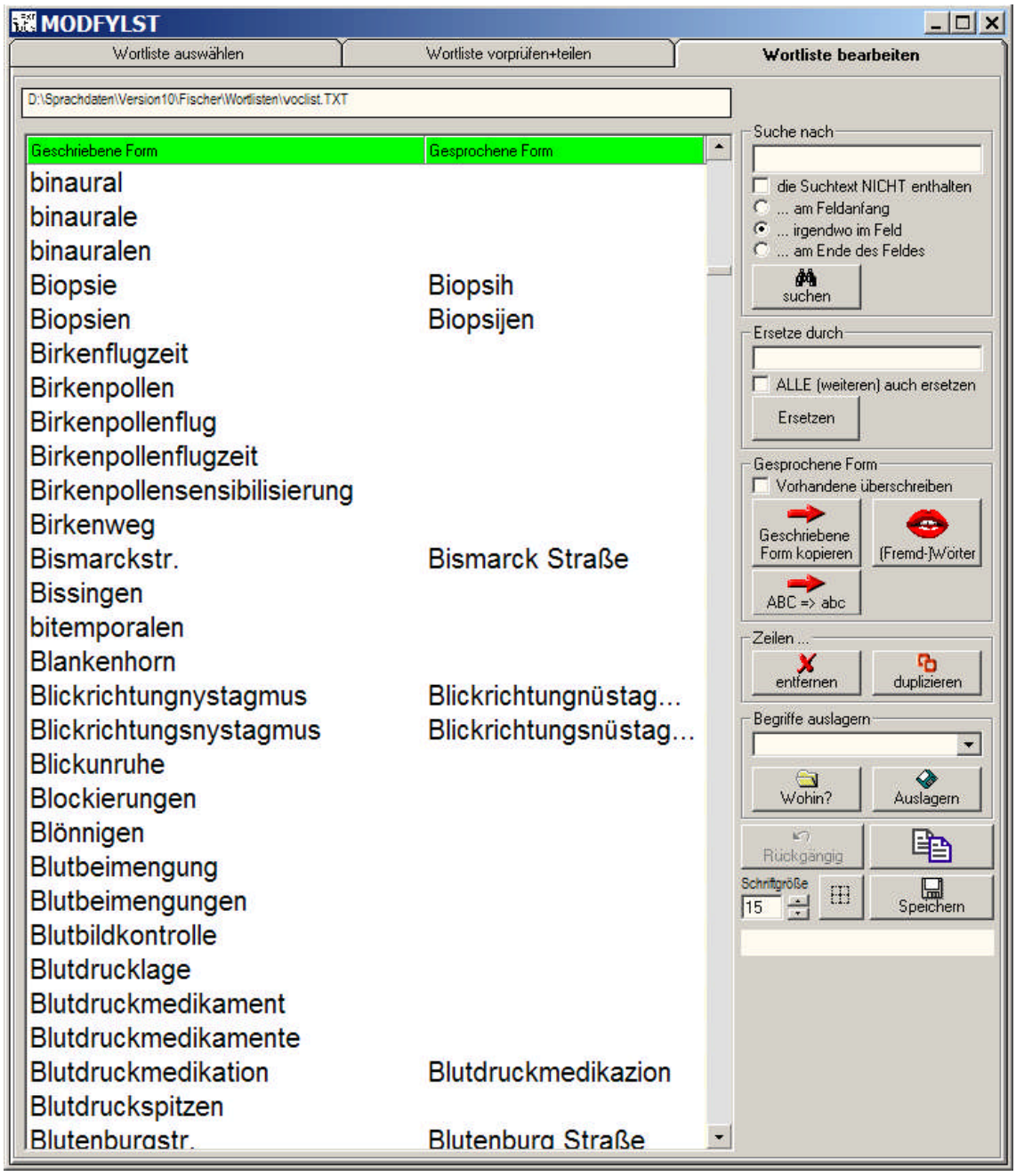
Falls man mehr als 1 Rechtschreibprüfung in MS Word implementiert hat, z.B. zusätzlich "Roche Rechtschreibprüfung Medizin", kann man noch festlegen, welches Programm im nächsten Schritt (s.u. "Wortliste vorprüfen+teilen") zuerst prüfen soll und welches als zweites und ob ab einem bestimmten Übereinstimmungsgrad eine automatische Ersetzung erfolgen soll. Bei Werten unter 100% sollte das Ergebnis sorgfältig kontrolliert werden. Automatische Fehlerbehebungen sind noch im Versuchsstadium.



Auf der Seite **Wortliste vorprüfen+teilen** kann man mittels der zugeschalteten Rechtschreibeprüfung von Microsoft Word und angehängten Wörterbüchern die Wortliste in richtig und falsch geschriebene (bzw. unbekannte) Begriffe teilen.



Der Kartenreiter **Wortliste bearbeiten** öffnet folgendes Fenster:

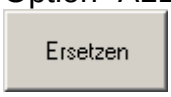


Es wird links die geschriebene Form und rechts die gesprochene Form (wenn abweichend) dargestellt. Die großen Farbbalken über den Spalten aktivieren (Farbe: grün) oder deaktivieren (Farbe: rot) die Spalte bezüglich Editierbarkeit.

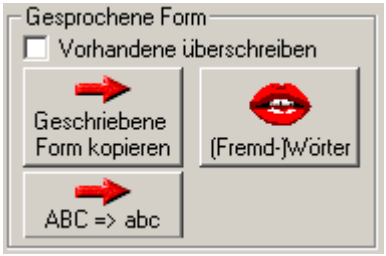
Suche nach
durchsucht die aktivierten Spalten nach dem eingegebenen Suchbegriff.
Option "NICHT": Markiert alle Felder, die den Suchbegriff nicht enthalten.

Löst den Suchvorgang aus.

Ersetze durch
ersetzt die markierten Bereiche durch den eingegebenen Ersetzungsbegriff.
Option "ALLE": Es werden alle markierten Bereiche ersetzt.



Löst den Ersetzungsvorgang aus.



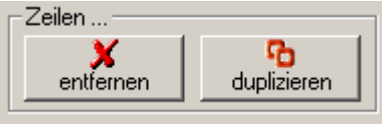
Gesprochene Form

Die Taste "**Geschriebene Form kopieren**" kopiert bei markierten Wörtern den Inhalt der geschriebenen Form (linkes Feld) in das Feld für die gesprochene Form (rechtes Feld) unter Berücksichtigung aller Tabellen des Tools **MODFYLST** außer Tabelle "(Fremd-)Wörter"! Danach kann man diese -falls nötig- manuell abändern, damit es der gesprochenen Form korrekt entspricht.

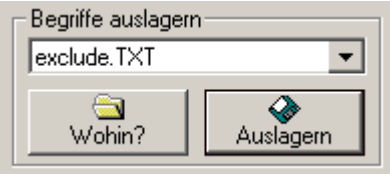
Bei Abkürzungen und Maßeinheiten wird tabellengesteuert die korrekte gesprochene Form sofort erzeugt. Zuerst listet man in einer separaten Tabelle alle Abkürzungen und Maßeinheiten in der geschriebenen und gesprochenen Form.
Hat man dann links z.B. die Maßeinheit "**l/m^2**" entsteht daraus für die gesprochene Form sofort "**Liter pro Quadratmeter"**. Gleiches gilt für Initialwörter:
für beispielsweise "**ADAC**" wird "**Ah Deh Ah Zeh**" erzeugt!

Die Taste "**(Fremd-)Wörter** kopiert bei markierten Wörtern den Inhalt der geschriebenen Form (linkes Feld) in das Feld für die gesprochene Form (rechtes Feld) unter Berücksichtigung aller Tabellen des Tools MODFYLST, d.h. auch der Tabelle "(Fremd-)Wörter. Sollte bereits eine gesprochene Form existieren, wird der Inhalt nur dann überschrieben, wenn im Feld "**Vorhandene überschreiben**" ein Häkchen gesetzt ist.

Die Taste "**ABC**" kopiert die links markierten Wörter 1:1 in Kleinbuchstaben nach rechts. So entsteht z.B. für den Begriff "ZYPREXA" die gesprochene Form "zyprexa".

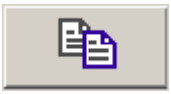


Mit den Taste "entfernen" bzw. "duplizieren" entfernt man oder dupliziert man markierte Zeilen.



Begriffe auslagern
Im Auswahlfeld werden Dateinamen aufgelistet, unter denen man bereits markierte Zeilen ausgelagert hat. Mit der Taste "Wohin?" gibt man dabei den Ort (d.h. Unterverzeichnis) an. Mit der Taste "Auslagern" startet man die Operation. "Auslagern" in diesem Zusammenhang heißt "Kopieren des gesamten Inhalts aller (teil-)markierten Zeilen in eine andere Datei."

Macht die letzte Änderung rückgängig.


Die Taste (einmal am Anfang EIN oder AUS) stellt jeden Begriff, der im linken Fenster angeklickt wird, automatisch in die Windows Zwischenablage. Wenn man nun parallel das TextPrep-Tool CONTEXT geöffnet hat, sieht man unmittelbar die Textumgebungen, in denen dieser Begriff vorkommt.

Schriftgröße
Auch hier kann man die Schriftgröße zur besseren Lesbarkeit nach den persönlichen Bedürfnissen anpassen.

20. DOC2TXT

Üblicherweise werden Quelldaten in einem MS Word-kompatiblen Format bereitgestellt. Da TextPrep nur mit dem eigentlichen Text ohne jegliche Formatierung arbeitet, muss man nahezu als allererstes die Quelldaten in das TXT-Format umsetzen. Dies lässt sicht leicht mit Hilfe des Tools DOC2TXT bewerkstelligen. Dieses Tool veranlasst MS Word, die Dateien mit MS Word-kompatiblen Dateierweiterungen **DOC, RTF, HTM** usw. aus dem Quellverzeichnis einzulesen und im TXT-Format in das Zielverzeichnis zu speichern. Die Größe der Dateien verkleinert sich auf weniger als 20 Prozent der ursprünglichen Größe.

Bei Dateien von MS Excel (Dateierweiterung **XLS**) werden die Feldinhalte ausgelesen. Jeder Feldinhalt wird in eine separate Zeile geschrieben und automatisch mit einem Satzende-Punkt versehen.
Leere Felder und Felder, die nur eine ganze Zahl, eine Dezimalzahl (z.B. einen Betrag) oder ein Datum enthalten, werden übersprungen.

PDF-Dateien können mit dem Tool DOC2TXT ebenfalls in das TXT-Format konvertiert werden.
Bei dieser Konversion treten zum ggw. Zeitpunkt gelegentlich Wortbrüche auf.
Z.B. ziemlich -> ziem lich
Sollte dies stören, ist die Verwendung des Programms Nuance PDF-Converter zu empfehlen.

Bei anderen Dateiformaten muss die Konversion in das TXT-Format ggf. manuell oder mit geeigneter Software vorgenommen werden.

Achtung:
Bei der Verwendung dieses Tools wird oft eine große Anzahl von Dateien konvertiert. Es ist daher ratsam, vor dem Starten des Tools den evtl. installierten Virenscanner zu deaktivieren. Manche Virenscanner erzeugen beim Einlesen so vieler Dateien einen Programm- bzw. einen Systemabsturz.
Nicht vergessen, den Virenscanner danach wieder zu aktivieren.!

21. SPACED

Dieses Tool erfasst alle gesperrt geschriebenen Ausdrücke des Quelltextes. Auf diese Weise kann man auch solche Wörter wieder mittels REMOVERR in ihren Normalzustand zurückführen und so bei der Erzeugung der Bi- und Trigrammstatistiken und beim Sprachmodell (Dragon VocTool) berücksichtigen.

Die gefundenen Begriffe werden in gesperrter Form in der Datei "SperrdruckWeit.txt" und in normaler Form in der Datei "SperrdruckEng.txt" gespeichert.

D) Anleitung zur Vorgehensweise

Es heißt im Sprichwort: **„Viele Wege führen nach Rom“.** Und so verhält es sich auch mit Anleitungen zur Vorgehensweise. Es soll hier beispielhaft aufgezeigt werden, wie man vorgehen kann. Es gibt viele Alternativen; in Abhängigkeit von der Qualität der vorliegenden Quelltexte können manche Schritte auch übersprungen werden.

Im Folgenden werden nun Hinweise gegeben, mit welchen Tools man zweckmäßigerweise welches Ergebnis erreicht, ohne die Schritte im Einzelnen zu erläutern, da diese im Abschnitt **Die Tools im Detail** dargestellt sind.

Es kann davon ausgegangen werden, dass auf die von Kunden zur Verfügung gestellten Texte folgende Eigenschaften (mehr oder weniger) zutreffen:

- sehr umfangreich
- enthält Tippfehler
- enthält Schreibfehler
- enthält sensitive, personenbezogene Daten von Dritten einschl. deren Namen und Anschrift
- inkonsistent bei der Verwendung von Abkürzungen und Maßeinheiten
- gemischt alte und neue deutsche Schreibweise

Nach vorliegender Erfahrung kann man wie folgt vorgehen:

1. Die Quelldateien in 1 Unterverzeichnis kopieren.
2. Schreibschutz bei manchen Dateien entfernen.
3. Umsetzen der Quelldateien in TXT-Dateien
4. Fremdsprachige Dateien entfernen.
5. Dateien umbenennen
6. Doppelte Dateien entfernen.
7. Umsetzen von ASCII-Umlauten (DOS) in ANSI (Windows) (falls erforderlich.
8. Entfernen aller Personennamen und -anschriften von Dritten, falls vom Auftraggeber gewünscht
9. Texte zu 1 Datei zusammenfassen. Leerzeichen und Tabs am Zeilenanfang entfernen, Leerzeichen am Zeilenende entfernen
10. Textumbrüche entfernen
11. Alle Anführungszeichen u. dgl. vereinheitlichen oder entfernen oder durch ein Leerzeichen ersetzen
12. Gesperrt geschriebene Begriffe in die Normalform zurückführen
13. Umsetzen der Texte in einheitliche, neue deutsche Schreibweise.
14. Entfernen von Tipp- und Schreibfehlern durch Bearbeiten temporär erzeugter Wortlisten
15. Erzeugen von endgültigen Wortlisten
16. Hinzufügen der gesprochenen Formen bei Abkürzunge, Maßeinheiten und besonderen Ausdrücken (z.B. „ADAC“)
17. Eingabe der Wortlisten in das Dragon VocTool und Erzeugen des Vokabulars.

In der nun folgenden Darstellung wird auf diese Einteilung 1. – 17. Bezug genommen.

1. Die Quelldateien in 1 Unterverzeichnis kopieren.

Dies kann man manuell oder durch ein geeignetes Tool erreichen. (Nicht Bestandteil von TextPrep).

2. Schreibschutz bei manchen Dateien entfernen.

Wenn festgestellt wird, dass sich manche Quelldateien einer Änderung widersetzen, muss man z.B. im Windows Explorer die Dateien über
rechte Maustaste / Eigenschaften / Attribute überprüfen, ob der Schreibschutz aktiviert ist. Diesen dann entfernen.

3. Umsetzen der Quelldateien in TXT-Dateien.

Üblicherweise haben die Daten des Kunden ein MS Word-Format (.doc oder .rtf). Alle Formate, die von MS Word gelesen werden können, überführt man mit dem TextPrep-Tool **DOC2TXT** in das TXT-Format. Andere Formate muss man ggf. von Hand bzw. mit anderen Hilfsmitteln in das TXT-Format überführen.

4. Fremdsprachige Dateien entfernen

Außer in geringem Umfang kann ein deutschsprachiges DNS-Vokabular keine fremdsprachigen Wörter enthalten. Texte, die weitgehend in einer fremden Sprache geschrieben sind, müssen entfernt werden.
Es empfiehlt sich, mit dem TextPrep-Tool CONTEXT nach Begriffen wie "the" oder "and" zu suchen, um englischsprachige Texte zu lokalisieren. Ähnlich muss man bei anderen Sprachen vorgehen. Die entsprechenden Texte bzw. Dateien sollten gelöscht werden.

5. Dateien umbenennen

Aus Sicherheitsgründen sollten die Namen der Quelltexte mit dem TextPrep-Tool RENAME umbenannt werden.

6. Doppelte Dateien entfernen

Überraschenderweise enthält der Quelltext sehr oft inhaltlich gleiche Dateien.
Es gibt unter den Freeware- und Sharewareangeboten gute Tools, die inhaltlich doppelten Dateien aufspüren und entfernen. (Nicht Bestandteil von TextPrep).

7. Umsetzen von ASCII-Zeichen in ANSI-Zeichen

Besonders deutsche Umlaute sind betroffen, wenn noch Texte aus der DOS-Ära vorliegen. Man erkennt dies leicht durch falsche Darstellung der deutschen Umlaute. Unter Verwendung von **SHOWHEX** schaut man sich stichprobenhaft die hexadezimalen Besonderheiten an, die farbig dargestellt sind.

Liegen tatsächlich einige Texte in ASCII-Format vor, separiert man diese Texte von den anderen durch **WRONGHEX** und ersetzt diese falschen Zeichen durch die „richtigen“ mit dem Tool **REPLACE.**

Für die deutschen Umlaute sind folgende Umsetzungen erforderlich (hexadezimal dargestellt):
94 -> F6 (ö), 84 -> E4 (ä), 81 -> FC (ü), E1 -> DF (ß), 9A -> DC .(Ü), 8E -> C4 (Ä), 99 -> D6 (Ö), F5 -> A7 (§).

8. Entfernen aller Personennamen

Entfernt werden nur Namen von Personen, wenn der Inhalt der Texte schutzwürdig in Bezug auf diese Personen ist. Es kann sehr sinnvoll sein, die Namen und Adressen von Geschäftskontakten **nicht** zu entfernen.

a) Umändern von Dateinamen

Oft lassen Dateinamen Rückschlüsse auf den Patienten zu, und sei es über eine Chiffre-Nummer. Diese Dateinamen lassen sich mit dem Tool **RENAME** mit einer Anweisung einheitlich umbenennen.

b) Streichen von Textabschnitten am Anfang der Dokumente

In aller Regel sind die Texte nach einer oder mehreren Vorlagen aufgebaut. Man ermittelt nun, welche Vorlagen häufig vorkommen und merkt sich Leitbegriffe. Es folgt nach Absender, Adresse und Name/Anschrift des Patienten z.B. in einer der nächsten Zeilen das Wort „Diagnose“. Dann separiert man alle Texte mit diesem Muster durch das Tool **TXTSTART** und löscht alle Texte vom Anfang der Dokumente bis zu dem Begriff „Diagnose“ durch das Tool **DELLINES**.

c) Streichen von Textabschnitten innerhalb der Dokumente

Bei manchen Berichten wird bei jeder neuen Seite die Seitenzahl angegeben, gefolgt von dem Patientennamen. Diese Zeilen, erkenntlich durch die besondere Art der Darstellung der Seitenzahl, lassen sich mit einer anderen Option des Tools **DELLINES** löschen.

d) Streichen von Textabschnitten am Ende der Dokumente

Vor allem bei wissenschaftlichen Veröffentlichungen schließen diese mit dem Kapitel „Referenzen“ bzw. „Literaturangaben“. Mit dem Tool **DELLINES** können auch diese Textabschnitte leicht entfernt werden.

e) Streichen/Ersetzen von Personenname im Fließtext

Damit der Satz in seinem Kontext nicht zu sehr gestört wird, ist es eher sinnvoll, die Namen von Personen durch Begriffe wie „der Patient“ oder „die Patientin“ zu ersetzen.
Dies lässt sich bewerkstelligen mit dem Tool **REPLACE**, insbesondere mit der **besonderen Ersetzungsanweisung [w]**.

9. Texte zu 1 Datei zusammenfassen

Texte lassen sich mit dem TextPrep-Tool **COMBFILE** zu 1 Datei zusammenfassen. Dies beschleunigt die nachfolgenden Such- und Korrekturoperationen. Bei COMBFILE folgende Optionen auswählen:

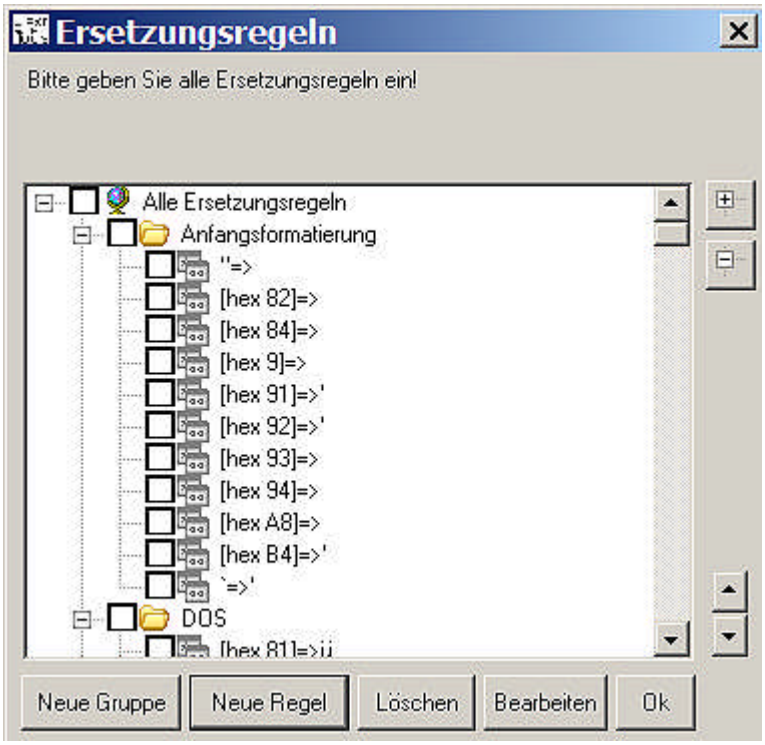
Leerzeichen und Tabs am Zeilenanfang entfernen
Leerzeichen am Zeilenende entfernen
Umbrüche durch Leerzeichen ersetzen

10. Textumbrüche entfernen

Textumbrüche entfernt man weitgehend mit dem TextPrep-Tool **NOFRAGMT**.

11. Alle Anführungszeichen u. dgl. vereinheitlichen oder entfernen oder durch ein Leerzeichen ersetzen

Es gibt z.B. ein- und zweigestrichenen Anführungszeichen, senkrechte oder schräge Auslassungszeichen usw. Man kann diese Zeichen mit dem TextPrep-Tool **REPLACE** vereinheitlichen bzw. löschen. Die erforderlichen Ersetzungsanweisungen in der zugehörigen Tabelle "Ersetzungsregeln" könnten wie folgt aussehen:



Bitte beachten, da es in der Abbildung nicht erkennbar ist:
Das Hex-Zeichen 09 wird durch ein Leerzeichen ersetzt.
Die Hex-Zeichen 82, 84, 93, 94 und A8 werden durch nichts ersetzt, d.h. gelöscht.

12. Gesperrt geschriebene Begriffe in die Normalform zurückführen

Das TextPrep-Tool **SPACED** erfasst diese Begriffe. Mit dem Tool **REMOVERR** und der durch SPACED erzeugten Datei "SperrdruckWeit.txt" werden diese Begriffe im Quelltext in ihre Normalform überführt.

13. Umsetzen der Texte in neue deutsche Rechtschreibung

Im Tool **REMOVERR** wird die mitgelieferte Datei „AltZuNeu.txt“ als Korrekturliste verwendet und auf die im Quellverzeichnis gespeicherten Texte angewendet. Dabei unbedingt die Option **Ersetzungen auch auf Teile von Wörtern anwenden** aktivieren.

14. Entfernen von Tipp- und Schreibfehlern

Es werden nacheinander temporäre Wortlisten (ohne gesprochene Form) erzeugt, und zwar durch das Dragon VocTool und durch die TextPrep-Tools **ABBREVNS** (Abkürzungen), **INITIALS** (Wörter mit mehr als 1 Großbuchstaben), **NOVOWELS** (Maßeinheiten), **PHRASES** (Ausdrücke wie Hals- und Beinbruch), **MULTTERM** (Bindestrichworte). Jede einzelne Liste wird nun mit dem Tool **REMOVERR** auf Schreib- und Tippfehler untersucht und korrigiert. Dabei ist es hilfreich, parallel das Tool **CONTEXT** zu starten, um für jeden Begriff sofort die Textumgebung angezeigt zu bekommen.

Technischer Hinweis: Hier ist das Arbeiten mit 2 Bildschirmen besonders hilfreich. Sollte die Grafikkarte den Anschluss von 2 Bildschirmen zulassen, ist unter Windows XP keine weitere Hard- oder Software erforderlich.

Am Ende jeder Liste überträgt dann REMOVERR alle Korrekturen auf den gesamten Quelltext.

Durch sorgfältiges Arbeiten in dieser Phase erhält man später praktisch fehlerlose Vokabulare und die zugehörigen Kontextstatistiken sind nicht durch fehlerhafte Worte beeinträchtigt.

15/16. Erzeugen von endgültigen Wortlisten, Hinzufügen der gesprochenen Form

Mit den Tools **PHRASES** und **MULTTERM** erzeugt man nun Wortlisten mit den gesprochenen Formen. Die Tools **ABBREVNS**, **INITIALS** und **NOVOWELS** erzeugen Wortlisten ohne gesprochene Form. Eine gesprochene Form bei diesen 3 Wortlisten ist jedoch unbedingt erforderlich. Diese übernimmt man aus früheren Wortlisten auf elegante Weise mit dem Tool **2LISTS**. Liegen keine gesprochenen Formen vor, kann man diese zeitsparend mit dem Tool **MODFYLST** erzeugen.

Mit den Tools **NEXTWORD** und **XPRESSNS** werden weitere Wortlisten erzeugt, die jedoch nur in geringem Umfang eine gesprochene Form erfordern.

17. Eingabe alle Wortlisten in das VocTool

Die fertigen Wortlisten mit ihren gesprochenen Formen gibt man schließlich an entsprechender Stelle in das Nuance Voctool ein und erzeugt so ein fehlerloses Vokabular.

>>> ENDE <<<