

Content Filter ###
#####

Idee:

- Content Filter: modifiziere Text so, dass er den Google Content Filter passiert und der modifizierte Text nicht als duplicated content angesehen wird - und der text für menschen trotzdem leserlich bleibt

Aktuelle Umsetzung:

- unterscheidung in engl und deutsche texte
- aktueller Release mit folgenden funktionen:
 - Number2Word(): wandele eine zahl in ein englisches wort um
 - PluralManipulation(): bilde den plural von englischen wörtern
 - StemmingManipulation(): reduzierung von englischen wörtern auf ihre grundform
- Zahl2Wort(): wandele eine zahl in ein deutsches wort um
- SynonymManipulation(): zufälliges ersetzen von deutschen wörtern mit einem zufälligem synonym zum ursprünglichen wort
- GrundformManipulation(): reduzierung von deutschen wörtern auf ihre grundform
- TextBlowup(): füge satzkombinationen zum array(=modifizierender text) hinzu -> nimm randomized einzelne wörter aus dem vorhandenen Text - sortiere dieses nach alphabetischer reihenfolge - und speichere das ergebnis ans (aktuell:Ende) des zu modifizierenden Textes
- StringManipulation(): buchstaben, zeichen, wörter spiegeln, vertauschen, löschen, ersetzen ...:
 - Erste Buchstaben groß/klein schreiben
 - Wort groß/klein schreiben
 - zufälliges Zeichen einfügen
 - zufällige Nummer einfügen
 - zufälligen Buchstaben einfügen
 - Wörter zufällig verdoppeln
 - Wörter zufällig verdoppeln und das verdoppelte spiegeln
 - Wörter zufällig spiegeln
 - zeichen aus einem Wort zufällig löschen
 - Deutsche Umlaute zufällig ersetzen
 - Wort zufällig mit unsinnigen zeichen füllen
- StringManipulation() ist mir noch zu krass, der text wird extrem unleserlich bis jetzt noch
 - src.txt: quellartikel harry potter
 - mod.txt: modifizierter harry potter artikel

Todos:

- StringManipulation(): einstellen, dass der text nicht zu unleserlich wird
- amazon content grabbing + adding: hole randomized amazonbeschreibung, modifiziere den content davon leicht und baue dieses in den zu modifizierenden Text ein
- statische synonym textdatei nehmen und diese dann abarbeiten, wenn kein treffer gefunden wurde dann benutze das langsamer my \$result =

Lingua::DE::Wortschatz::use_service('T', \$_);

- hauptkeyword(s) des textes müssen von jeglicher modifikation ausgeschlossen werden
- den text nach allen modifikationen wieder in eine leserliche form bringen: ergo 2x \n nach hauptkeyword und 1 x \n nach jeweils 5-7 satzenden (sprich nach 5-7 ". oder ! oder ?")
- later:

hauptkeywords mit mod_rewrite suchanfragen verknüpfen
code qualität steigern (oo-style ; use strict; etc)

Einsetzbar:

- zoozle
- p2p-blog
- rb3
- generel seo/sem (darum, darf das programm nicht in andere/falsche hände fallen)

Konzept:

- ??? (Torsten)

Realisierung:

- Basti