##########################
Content Transformer
##########################

Idee:

- Content Filter: modifiziere Text so, dass er den Google Content Filter passiert und der modifizierte Text nicht als dublicated 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
- StemmmingManipulation(): 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 vedoppelte 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 langsame 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:

Grundsätze:

- Ergebniss muss so aussehen, als ob es ein Mensch geschreiben hätte
    - Keine StemmmingManipulation():
    - Keine StringManipulation()
    - TextBlowup() muss rekonzeptioniert werden

How to do:

- Focus af Synonyme legen
    - Synonymdatenbanken nutzen
    - Wortstämme erkennen
        - Bsp.: "biligual" -> "zweisprachig" also wird "biliguale" -> "zweisprachige"
    - Groß/Kleinschreibung beachten
    - "Denglisch" erkennen
- Antonyme
    - Verwenden, wenn kein Synonym vorhanden
        - Antonyme verneinen
            - z.B. Antonym von "gut" wäre "schlecht" -> "gut" wird also ersetzt durch "nicht schlecht"
- Formatierung
    - Ersetze Synonyme mit html zu 50% formatieren (, <b>, <i>, <a> ..)
    - Leerzeilen nach Sätzen zufällig einfügen (1 Leerzeile pro Artikel)
    - Überschriften mit hx formatieren
    - Zufallsberechnung: den ersten Satz oder Wortgruppe als <b> oder formatieren
- Grammatik:
    - Sätze mit 2 Kommas als Einschübe
        - z.B. "..., hier steht text, ..." wird "...- hier steht text - ...."
- Datum ist Datum des Quellcontents minus 1 Tag.

Realisierung:

- Basti

Synnonymdatenbanken:

http://www.wie-sagt-man-noch.de/synonyme/ -> Note 3,3
http://www.wie-sagt-man-noch.de/synonymliste/ -> spidern für eingene DB?

http://wortschatz.uni-leipzig.de/ -> Note 2,3
http://www.woerterbuch.info/ -> Note 3,0
http://synonyme.woxikon.de/synonyme/ -> Note 3,0
http://elsap1.unicaen.fr/cgi-bin/cherches.cgi -> französisch
http://www.woerterfinden.de/ -> gut für Substantive