

Technisches Konzept	
Texte leistungsschutzgerecht und rechtlich sicher verändern	
Dokumentname: „Technisches-Konzept.docx“	Versionsnummer: 1.2.2
Erstellt am: 05.10.2015 Sang Tanzer	Geprüft am: 15.10.2015 Sebastian Enger / Position

Inhalt

1. Technische Idee	2
2. Funktionalität	2
2.1. Als Online Tool	2
2.2. Als Plugin in bestehender Schreib-/Dokumentensoftware	2
3. Oberfläche	3
4. Anwendung	6
5. Topics zu bearbeiten	6
5.1. Tool	6
5.2. Sprachzeichen	6
5.3. Rechtschreibung	6
5.4. Synonymen	7
5.5. Anglizismen	7
5.6. Grammatikalischen Satzstellungen	7
5.7. Akronymen	7
5.8. Paraphrasierungen	8
6. Ablauf	8
6.1. Informationssammlung	8
6.2. Software Prototyp	8

1. Technische Idee

Textinhalte datenbankgestützt bei gleichbleibender Bedeutung anpassen und optimieren.

2. Funktionalität

2.1. Als Online Tool

Primär wird die Nutzung in der Onlineumgebung realisiert.

1. Bei OneTipp kann man einfach seinen geschriebenen Text reinkopieren
2. Mit Auslösen des Senden Buttons erhält man zu seinem geschriebenen Text alternative Textzusammenstellungen, die inhaltlich das gleiche Ergebnis erzielen, jedoch mit
 - a. Paraphrasierungen
 - b. Synonymen
 - c. Akronymen
 - d. Anglizismen
 - e. Grammatikalischen Satzumstellungen
3. Diesen Text kann man anschließend in seinen Blog oder auf Social Media Kanälen verbreiten
4. **Ausgegebene Texte sind Leistungsschutzkonform und verletzen keine Rechte möglicher Urheber mehr**

2.2. Als Plugin in bestehender Schreib-/Dokumentensoftware

Sekundär wird das leistungsfähige Tool als Plugin für Desktop- bzw. Mobileanwendungen bereit gestellt.

5. Das Plugin kann sowohl in Officeprogrammen online wie offline wie auch in mobilen Textbearbeitungsanwendungen eingesetzt werden.
6. Die Funktionalität wird genau gleich sein, wie beim Onlinetool. Nur der Konvertierungsverweis wird in der Navigationsstruktur des jeweiligen Programmes eingebunden sein.
7. Mit Auslösen des KONvertierungs-Buttons erhält man zu seinem geschriebenen Text ebenso alternative Textzusammenstellungen, die inhaltlich das gleiche Ergebnis erzielen, jedoch mit
 - a. Paraphrasierungen
 - b. Synonyme
 - c. Akronyme
 - d. Anglizismen
 - e. Grammatikalischen Satzumstellungen
8. Dies ermöglicht es dem Schreiber ein bestmögliches Ergebnis aus seinem erstellten Text heraus arbeiten zu können.

3. Oberfläche

Texteingabe

Reset Button ->

Erst gestern sorgte Seehofer mit einer spitzen Äußerung für Furore. Der bayrische Ministerpräsident drohte der Bundesregierung mit „wirksamer Notwehr“.

Sollte die Bundesregierung weiterhin für keine Begrenzung der Flüchtlingszahlen sorgen, müsse der Freistaat überlegen, was er mache, sagte der CSU-Chef am Mittwoch anlässlich eines Gesprächs mit den bayerischen Landräten und Oberbürgermeistern in Ingolstadt.

Die „Notwehr“-Äußerung machte prompt Schlagzeilen – die halbe Nation fragte sich: Was meint Seehofer damit?

Suchen Button ->

":

<p>Original Text</p> <p>Erst gestern sorgte Seehofer mit einer spitzen Äußerung für Furore. Der bayrische Ministerpräsident drohte der Bundesregierung mit „wirksamer Notwehr“. Sollte die Bundesregierung weiterhin für keine Begrenzung der Flüchtlingszahlen sorgen, müsse der Freistaat überlegen, was er mache, sagte der CSU-Chef am Mittwoch anlässlich eines Gesprächs mit den bayerischen Landräten und Oberbürgermeistern in Ingolstadt. Die „Notwehr“-Äußerung machte prompt Schlagzeilen – die halbe Nation fragte sich: Was meint Seehofer damit?</p>	<p>Modified Text</p> <p>zuerst gestern sorgte Seehofer mit einer spitzen Äußerung für Furore. . Der bayrische Ministerpräsident drohte der Bundesregierung mit „wirksamer Wildnis . Sollte die Bundesregierung weiterhin für keine Grenzlinie der Flüchtlingszahlen sorgen, . müsse der Freistaat rotseln , was er mache, . sagte der CSU-Chef am Mittwoch anlässlich eines Gesprächs mit den bayerischen Landräten und Oberbürgermeistern in Ingolstadt . Die „Notwehr“-Äußerung machte prompt Schlagzeilen – die halbe Nation fragte sich: Was meint Seehofer damit? ?</p>
---	--

TexteingabeReset Button ->

Patrick Herrmann hat sich einen Riss des hinteren Kreuzbandes im linken Knie zugezogen. Das ergaben intensive Untersuchungen bei Borussia's Mannschaftsarzt Dr. Stefan Hertl. Wegen anhaltender Kniegelenksbeschwerden wurden bei dem 24-Jährigen noch einmal intensive Behandlungen und Untersuchungen bei Mannschaftsarzt durchgeführt, wie die Fohlen auf ihrer Homepage berichten.

Aron Johannsson leidet an einer Nerven-Irritation im Bereich der rechten Hüfte. Diese Diagnose ergaben eingehende Untersuchungen bei Leistungsspezialist Dr. Jens Krüger in Berlin. Der US-Nationalspieler war am Mittwochmittag in die Hauptstadt gereist um Klarheit über die Verletzung zu bekommen. Seit Donnerstagnachmittag ist Johannsson zurück in Bremen. Um den betroffenen Nerv zu beruhigen wird der Stürmer in der Hansestadt konservativ behandelt. „Wir hoffen, dass Aron durch diese Art der Behandlung zeitnah wieder spielfähig ist“, so Cheftrainer Viktor Skripnik.

Suchen Button ->

DEBUG: Reading in Text from Filesystem

DEBUG: Parsing Satzzeichen

DEBUG: Preparing final Array for Working with

DEBUG: Doing parallel Tasks

DEBUG: Highlight Texts from Exchange

DEBUG: Writing final Text to FileSystem

";

Original Text

Patrick Herrmann hat sich einen Riss des hinteren Kreuzbandes im linken Knie zugezogen. Das ergaben intensive Untersuchungen bei Borussia's Mannschaftsarzt Dr. **Stefan Hertl**. Wegen anhaltender Kniegelenksbeschwerden wurden bei dem 24-Jährigen noch einmal intensive Behandlungen und Untersuchungen bei Mannschaftsarzt durchgeführt, wie die **Fohlen** auf ihrer Homepage berichten. Aron Johannsson leidet an einer Nerven-Irritation im Bereich der rechten Hüfte. Diese Diagnose ergaben eingehende Untersuchungen bei Leistungsspezialist Dr. Jens Krüger in Berlin. Der US-Nationalspieler war am Mittwochmittag in die **Hauptstadt** gereist um Klarheit über die Verletzung zu bekommen. Seit Donnerstagnachmittag ist Johannsson zurück in Bremen. Um den betroffenen Nerv zu beruhigen wird der Stürmer in der Hansestadt konservativ behandelt. „Wir hoffen, dass Aron durch diese Art der Behandlung zeitnah wieder spielfähig ist“, so Cheftrainer Viktor Skripnik.

Modified Text

Patrick **Haren (Ems)** hat sich einen Riss des hinteren Kreuzbandes im linken Knie zugezogen. Das ergaben intensive Untersuchungen bei Borussia's Mannschaftsarzt **D. abdicthen** Hertl. **durch** anhaltender Kniegelenksbeschwerden wurden bei dem 24-Jährigen noch einmal intensive Behandlungen und Untersuchungen bei Mannschaftsarzt durchgeführt, wie die **bestormen** auf ihrer Homepage berichten. **Ausd** Johannsson leidet an einer Nerven-Irritation im Bereich der rechten Hüfte. Diese Diagnose ergaben eingehende Untersuchungen bei Leistungsspezialist Dr. Jens Krüger in Berlin. Der US-Nationalspieler war am Mittwochmittag in die **Kraft** gereist um **Durchsichtigkeit** über die **Verstömmelung** zu bekommen. **seither** Donnerstagnachmittag ist Johannsson zurück in **-brunn**. Um den betroffenen **nerv** zu beruhigen wird der Stürmer in der **Gesichtspunkt** konservativ behandelt. „Wir hoffen, dass **Ausd** durch diese Art der **Heilverfahren** zeitnah wieder spielfähig ist“, so Cheftrainer Viktor **Schreibstil**.

Leipzig. Die Legida-Abspaltung „Offensive für Deutschland“ (OfD) will nun ebenfalls im wöchentlichen Rhythmus aufmarschieren. Nach Angaben der Leipziger Stadtverwaltung haben die Rechtspopulisten um Silvio Rösler am 17. und 31. Oktober Demonstrationen in der Messestadt angemeldet. Eine weitere Veranstaltung soll am Samstag dazwischen, am 24. Oktober, in Markkleeberg stattfinden. Zusammen mit den montäglichen Legida-„Spaziergängen“ muss die Leipziger Polizei somit nun jede Woche zwei rechte Aufmärsche absichern.

Erstmals war die „Offensive für Deutschland“ am 26. September auf dem Leipziger Augustusplatz in Erscheinung getreten. Unter den damals 350 Teilnehmern befanden sich auch zahlreiche Hooligans, die zum Teil gewaltbereit auftraten. Rings um den Aufmarsch versammelten sich damals Tausende Gegendemonstranten, protestierten vorwiegend friedlich, vereinzelt flogen aber auch Steine und die Polizei griff hart durch.

Suchen Button ->

DEBUG: Reading in Text from FileSystem
 DEBUG: Preparing Satzzeichen
 DEBUG: Preparing final Array for Working with
 DEBUG: Doing parallel Tasks
 DEBUG: Highlight Texts from Exchange
 DEBUG: Writing final Text to FileSystem

":

Original Text

Leipzig. Die Legida-Abspaltung „Offensive für Deutschland“ (OfD) will nun ebenfalls im wöchentlichen Rhythmus aufmarschieren. Nach Angaben der Leipziger Stadtverwaltung haben die Rechtspopulisten um Silvio Rösler am 17. und 31. Oktober Demonstrationen in der Messestadt angemeldet. Eine weitere Veranstaltung soll am Samstag dazwischen, am 24. Oktober, in Markkleeberg stattfinden. Zusammen mit den montäglichen Legida-„Spaziergängen“ muss die Leipziger Polizei somit nun jede Woche zwei rechte Aufmärsche absichern. Erstmals war die „Offensive für Deutschland“ am 26. September auf dem Leipziger Augustusplatz in Erscheinung getreten. Unter den damals 350 Teilnehmern befanden sich auch zahlreiche Hooligans, die zum Teil gewaltbereit auftraten. Rings um den Aufmarsch versammelten sich damals Tausende Gegendemonstranten, protestierten vorwiegend friedlich, vereinzelt flogen aber auch Steine und die Polizei griff hart durch.

Modified Text

Sami. Die Legida-Abspaltung „Offensive für **Finken** (OfD)) will nun ebenfalls im wöchentlichen Rhythmus aufmarschieren. Nach **Daten** der Leipziger Stadtverwaltung haben die Rechtspopulisten um **umgangssprachlich** Rösler am 17. und 31. **Gilbhart** Demonstrationen in der **Arbeitsscheu** angemeldet. Eine weitere **Anlass** soll am Samstag dazwischen, am 24. **Silbhart**, in Markkleeberg stattfinden. **gemeinsam** mit den montäglichen Legida-„Spaziergängen“ muss die Leipziger **Erdaltertum** somit nun jede **sieben Tage** zwei rechte Aufmärsche absichern. **erstmal**s war die „Offensive für **Finken** am 26. **Neunte** auf dem Leipziger Augustusplatz in Erscheinung getreten. Unter den damals 350 Teilnehmern befanden sich auch zahlreiche Hooligans, die zum Teil gewaltbereit auftraten. Rings um den Aufmarsch versammelten sich damals Tausende Gegendemonstranten, protestierten vorwiegend friedlich, vereinzelt flogen aber auch Steine und die **Erdaltertum** griff hart durch.

"

Original Text

Leipzig. Die Legida-Abspaltung „Offensive für Deutschland“ (OfD) will nun ebenfalls im wöchentlichen Rhythmus aufmarschieren. Nach Angaben der Leipziger Stadtverwaltung haben die Rechtspopulisten um Silvio Rösler am 17. und 31. Oktober Demonstrationen in der Messestadt angemeldet. Eine weitere Veranstaltung soll am Samstag dazwischen, am 24. Oktober, in Markkleeberg stattfinden. Zusammen mit den montäglichen Legida-„Spaziergängen“ muss die Leipziger Polizei somit nun jede Woche zwei rechte Aufmärsche absichern. Erstmals war die „Offensive für Deutschland“ am 26. September auf dem Leipziger Augustusplatz in Erscheinung getreten. Unter den damals 350 Teilnehmern befanden sich auch zahlreiche Hooligans, die zum Teil gewaltbereit auftraten. Rings um den Aufmarsch versammelten sich damals Tausende Gegendemonstranten, protestierten vorwiegend friedlich, vereinzelt flogen aber auch Steine und die Polizei griff hart durch. Im Gegensatz zum Auftakt an prominenter Stelle in der Leipziger City zieht es Rösler und Co. nun offenbar in den Stadtteile. Ihre nächste Veranstaltung soll nach eigenen Angaben am 17. Oktober um 17 Uhr am Grünauer Allee-Center stattfinden. Am Samstag danach (24. Oktober) soll es am Rathaus in Markkleeberg weitergehen und für den 31. Oktober planen die Rechtspopulisten einen Aufmarsch am Nordplatz im Leipziger Stadtteil Gohlis. Sowohl in Leipzig, als auch in Markkleeberg sind diverse Gegenveranstaltungen in Vorbereitung, hieß es am Donnerstag. Genaue Angaben zu allen Demonstrationen waren am Donnerstag noch nicht aus den zuständigen Ordnungsämtern zu erfahren. Bis zum Sommer war Silvio Rösler noch Vordenker bei Legida. Danach trennten sich die Wege und die OfD trat als weitere rechtspopulistische Initiative in Erscheinung. Unterstützt wird die Gruppe auch vom GIDA-Dachverband, der bisher rechte Aufmärsche im Leipziger Umland organisiert hat.

Modified Text

Samisch. Die Legida-Abspaltung „Offensive für **Puschen** (OfD)) **will** nun ebenfalls im wöchentlichen Rhythmus aufmarschieren. Nach Angaben der Leipziger **Gemeindehaus** haben die Rechtspopulisten um **umgangssprachlich** Rösler am 17. und 31. **Rosenkranzmonat** Demonstrationen in der Messestadt angemeldet. Eine weitere **Fest** soll am **Sommerschirm** dazwischen, am 24. Oktober, in Markkleeberg stattfinden. **kooperativ** mit den montäglichen Legida-„Spaziergängen“ muss die Leipziger **Pal♠ozoikum** somit nun jede **sieben Tage** zwei rechte Aufmärsche absichern. Erstmals war die „Offensive für **Puschen** am 26. **Scheidung** auf dem Leipziger Augustusplatz in **Oberfl♠che** getreten. Unter den damals 350 Teilnehmern befanden sich auch zahlreiche Hooligans, die zum Teil gewaltbereit auftraten. Rings um den **Heerschau** versammelten sich damals Tausende Gegendemonstranten, protestierten vorwiegend friedlich, vereinzelt flogen aber auch Steine und die **Pal♠ozoikum** griff hart durch. Im **indes** zum Auftakt an prominenter Stelle in der Leipziger City zieht es Rösler und Co. nun offenbar in den Stadtteile. Ihre nächste **Fest** soll nach eigenen **Unterlagen** am 17. **Rosenkranzmonat** um 17 Uhr am Grünauer Allee-Center stattfinden. Am **Sommerschirm** danach (24. Oktober) soll es am Rathaus in Markkleeberg weitergehen und für den 31. **Rosenkranzmonat** planen die Rechtspopulisten einen **Heerschau** am Nordplatz im Leipziger **Ma♠ Gohlis**. Sowohl in **Samisch**, als auch in Markkleeberg sind diverse Gegenveranstaltungen in **♠bung**, hieß es am Donnerstag. Genaue **Unterlagen** zu allen Demonstrationen waren am Donnerstag noch nicht aus den zuständigen Ordnungsämtern zu erfahren. Bis zum Sommer war **umgangssprachlich** Rösler noch Vordenker bei Legida. Danach trennten sich die Wege und die OfD trat als weitere rechtspopulistische Initiative in **Oberfl♠che**. Unterstützt wird die **Typ** auch vom GIDA-Dachverband, der bisher rechte Aufmärsche im Leipziger Umland organisiert hat.

4. Anwendung

5. Topics zu bearbeiten

5.1. Tool

- Textzeichen am Anfang vom „Original Text“ entfernen
- Bei reinkopierten Absätzen sollten im Ergebnis auch einen Absatz einziehen (Schönheitskorrekturen sind erst zum Ende wichtig, zuerst Funktionalität)

5.2. Sprachzeichen

- Umlaute richtig ausgeben

5.3. Rechtschreibung

- Eigennamen sollten abgeprüft werden. Unter anderem anhand von zusätzlichen Kategoriewörtern, die wir in der Datenbank als feste Datenbankeinträge/-listen erfassen.

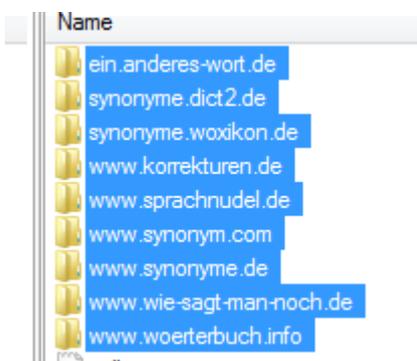
<https://de.wikipedia.org/wiki/Eigennamen>

- o Beispiel *Ort*: in Untergruppenbach, Neu Lübbenau
- o Beispiel *Land*: in Nordkorea, in Afghanistan am Hindukusch
- o Beispiel *Fluss*: an der oberen Donau
- o Beispiel *Verein*: der Verein Borussia Mönchengladbach
- o Beispiel *Unternehmen*: das Unternehmen Beiersdorf
- o Beispiel *Institutionen*: Universität Leipzig, Bernd-Blindow-Schule
- o Beispiel *Produkte*: Ferrero Rocher

- Beispiel *Tiernamen*: der Tausendfüßler
- Personennamen: Herr Winterkorn
- Namenskombination: Martin Winterkorn (Abprüfen eines Vor- & Nachnamens)
- Anmerkung SE:
 - Scanne aktuell (menschliche) Vor und Nachnamen -> Aufbau einer eigenen Namensdatenbank:
 - <http://www.onomastik.com/>
 - <http://www.firstname.de/>
 - <http://www.namepedia.org/>
 - <http://www.beliebte-vornamen.de/>
 - <http://www.vornamenarchiv.de/>
 - <http://www.vorname.com/>
- Eventuell hier Code hernehmen:
 - <http://demo2-opener.rhcloud.com/welcome.action>
 - <http://www.opener-project.eu/project/demos/>
- Priorisierung:
 1. Vor und Nachnamen
 2. Unternehmen / Institute
 3. Produkte
 4. Ort und Länder

5.4. Synonymen

Aktuelle Quellen:



5.5. Anglizismen

5.6. Grammatikalischen Satzumstellungen

- Abfrage der Zeitform
- Präpositionalphrasen
 - Präposition + Nominalphrase: vor dem dunklen Stuhl, seit einer Stunde
 - Präposition + Adverb: seit damals, vor morgen

5.7. Akronymen

TODO: Recherche, welche Webseiten eventuell nur Akronyme haben

5.8. Paraphrasierungen

Die Paraphrasierungen gehören in die komplexere Strukturierung der Computer Linguistik.

6. Ablauf

6.1. Informationssammlung

- Recherche welche Tools vorhanden sind
- Welche frei verfügbaren Informationsquellen können wir kostenlos anzapfen

6.2. Software Prototyp

- Aufbau einer Synonymdatenbank
- Aufbau einer Namensdatenbank
- Aufbau einer Akronymdatenbank
 - => freie Informationsquellen nutzen
 - „Alles erledigen was erstmal geht, ohne einen Computerlinguisten“
- Einbau der SYN, Namens, Akronym DB
- Kontinuierliches Testing von Sang und Basti
- Software Prototyp 1:
 - Funktionalität:
 1. Erkennung von Vor und Nachnamen
 2. Ersetzen von Synonymen und Akronymen
 3. Weboberfläche
 4. Massive Steigerung der Ergebnisqualität:
 - gross geschriebene (Such)Worte nur mit gross geschrieben synonymen austauschen
 - Datenbank verbessern -> statt ein „Startwort“ und dazu Synonyme ganz einfach Startwort+Synonyme in eine Tabelle schreiben und darüber mit Sphinx suchen lassen -> dann beim Sphinx Suchen noch gute Filter mit einbauen
- Software Prototyp 2:
 - Stabiler, bugfreier Prototyp 1 mit Basisfunktionalität
 - Bereinigte Synonym- und Namensdatenbank (Keine Dubletten mehr)
 1. QA auf die Namens und Synonym Datenbank
 - Sumy Textzusammenfassung einbauen
 - Austausch Synonyme nicht doppelt verwenden
 - SynRankCalculator: Berechne, wie gut ein Synonym passend ist
 1. Simple: Synonymwortlänge, enthält Sonderzeichen, enthält Leerzeichen oder besteht aus mehreren Worten -> dann Wert 1 – Penalty
 2. Später dann bei der Auswahl des Synonyms das mit dem verbleibenden höchsten Rang verwenden

~~3. Stopwordliste definieren: wenn Stopword aus Liste in den Synonymen vorkommt, dann gib ordentlich Minuspunkte (erstmal unwichtig)~~

- Software Prototyp 3:
 - Grundlegende API Call Design fertig stellen
 - Speed & Schönheitsfehler („Absätze fehlen“) kümmern
 - Umlaute und Ausgabe von Schreibsonderlauten korrekt (kein String Ersetzen HACK)
 - ROCKSDB Datenbank als IN MEMORY Lösung (=Datenbank im Speicher halten für FAST ACCESS) - NamesDB als RocksDB im Hauptspeicher halten für schnelleren Zugriff
 - Web API (Wochenende: 17-18.10.2015)
 - Städtenamen, Monate, Flüsse etc als Datenbank mit aufnehmen (WO HERBEKOMMEN?) -> WikiPedia klauen
 - Wortschatz Leipzig PythON: <https://github.com/lehmannro/libleipzig-python> einbauen :
 1. Nutze zuerst Syn Leipzig
 2. Berechne SynRank (=python)
 3. Wenn SynRank == 1 oder >0.86 dann nimm SYN Leipzig
 4. Speichere Rawdaten von Leipzig in DB für spätere Aufbereitung
 5. Wenn SynRank(von Leipzig) kleiner als 0.85 dann nimm lokale SYN MYSQL DB
 6. Berechne SynRank von Lokale Mysql DB
 - Wenn SynRank(Leipzig) grösser oder gleich SynRank(Mysql): Nimm Syn Leipzig (lemminized Version)
 - Sonst: nimm Syn Mysql
 - Text Lesbarkeitsstatistiken: <https://pypi.python.org/pypi/textstat/>
 - Ergebniscodierungen: Welche Elemente wurden wie geändert?

Software Prototyp X:

- MUST HAVE: Transliteration („Übertragung von einem Sprachzeichensatz (zb. Latin-1) in einen anderen (zb. Kyrillisch, Simple Chinese, Arabic)
 - <https://pypi.python.org/pypi/transliteration>
 - <https://pypi.python.org/pypi/transliterate>
 - https://www.tablix.org/~avian/blog/archives/2009/01/unicode_transliteration_in_python/
 - <https://pypi.python.org/pypi/Unidecode>

Webschnittstelle:

- Definition Basics 1. Version: Wochenende 17/18.10.2015:
 - - post request – http POST
 - – customerid –Welchem Account dürfen wir die Anfrage in Rechnung stellen?
 - – textin – Einzugebenden Text
 - – resultcodes – werden immer mit ausgegeben: Siehe in Etwa hier:
 - http://wiki.melissadata.com/index.php?title=Result_Codes
 -
 - - transliteration enabled or not
 - - output: xml