

Verteilte Bearbeitung von Wortlisten.

**Entwurf Version 1
Anforderungsspezifikation
(Lastenheft)**

Mathias Nater

Juli 2008

Trennmuster-Opensource

Inhaltsverzeichnis

1	Ausgangssituation und Zielsetzung	3
2	Produkteinsatz	3
3	Produktfunktionen	4
3.1	Funktionale Anforderungen	4
3.1.1	Bereiche	4
3.1.2	Produktdaten	4
3.1.3	Produktleistungen	5
3.2	Nicht funktionale Anforderungen	5
3.2.1	Benutzbarkeit	5
3.2.2	Bedienbarkeit	5
3.2.3	Zuverlässigkeit	5
3.2.4	Effizienz	5
3.2.5	Änderbarkeit, Wartbarkeit	6
3.2.6	Übertragbarkeit	6
4	Entwicklungszyklus	6
5	Use-Case	7
6	Lieferumfang	7
7	Abnahmekriterien	7

1 Ausgangssituation und Zielsetzung

Die freie „Gruppe trennmuster.org“ stellt zwei Produkte her:

Wortliste Die Wortliste umfasst die häufigsten Wörter der deutschen Sprache inklusive deren Beugungen. Sie basiert auf einer älteren Liste von Werner Lemberg und wird laufend erweitert, indem weitere Listen eingepflegt werden. Zur Zeit (Stand Juli 2008) umfasst sie rund 500'000 Wörter und Wortformen.

Die Wortliste soll eine hohe Qualität bezüglich Rechtschreibung aufweisen. Sämtliche Wörter und Wortformen sind mit Silbentrennstellen versehen.

Neben der aktuellen Rechtschreibung (nach August 2006) werden auch die alte Rechtschreibung, sowie länderspezifische Besonderheiten berücksichtigt.

Die Wortliste ist derzeit auf <http://repo.or.cz/w/wortliste.git> gehostet.

Trennmuster Aus der Wortliste werden Trennmuster zur Weiterverwendung mit Liangs Silbentrennalgorithmus berechnet. Diese Trennmuster finden Eingang in die TeX-Live-Distribution und können auch von anderen Anwendungen wie OpenOffice verwendet werden.

Die Gruppe kommuniziert massgeblich über eine Google-Groups-Mailingliste (<http://groups.google.de/group/trennmuster-opensource>) und kennt keine festen Strukturen.

Es wird damit gerechnet, dass im Laufe der Zeit weitere Wortlisten zu Verfügung gestellt werden, die in die bestehende Wortliste eingepflegt werden sollen. Diese Arbeit ist mit grossem Aufwand verbunden und lässt sich kaum automatisieren.

Es wird deshalb angestrebt, eine Lösung für das verteilte manuelle Bearbeiten solcher Wortlisten zu finden. Dieses Dokument umfasst als Lastenheft die Anforderungen, die die Gruppe an eine solche Lösung stellt.

2 Produkteinsatz

Beim Produkt handelt es sich um ein HTML/CSS/JavaScript-Frontend, das mit einem Backend gekoppelt ist. Zur Implementierung des Backends werden keine Angaben zu Sprache/Technik gemacht.

Das Produkt wird verwendet um Wörter aus einer neuen Wortliste aufzubereiten (Rechtschreibung kontrollieren, Trennstellen einfügen, alternative Formen hinzufügen) und in die bestehende Wortliste einzupflegen.

Mit dem Produkt arbeiten vor allem die Mitglieder der Gruppe. Um grosse Mengen an Wörtern zu trennen, werden auch externe Arbeitskräfte hinzugezogen.

Das Produkt wird auf sehr heterogenen Betriebsumgebungen verwendet.

3 Produktfunktionen

3.1 Funktionale Anforderungen

3.1.1 Bereiche

Das Produkt umfasst einen geschützten Verwaltungsbereich und einen öffentlichen Betriebsbereich. Benutzer sind entweder dem Verwaltungs- und Betriebsbereich oder nur dem Betriebsbereich zugeordnet.

Innerhalb des Produktes gibt es keinen Bereich um dieses zu entwickeln oder zu verändern.

Im Verwaltungsbereich

- ... werden Benutzer verwaltet.
- ... werden Wörter zur Bearbeitung eingegeben.
- ... werden die Resultate der Bearbeitung abgeholt
- ... werden Informationen zu laufenden Betrieb angezeigt

Im Betriebsbereich

- ... werden Wörter zur Bearbeitung ausgegeben
- ... werden bearbeitete Wörter eingegeben
- ... werden bestimmte Informationen zum laufenden Betrieb angezeigt

3.1.2 Produktdaten

Daten werden nicht-destruktiv behandelt. Das heisst zum Beispiel, dass ein Wort aus einer Wortliste kopiert, bearbeitet und in einer neuen Liste mit bearbeiteten Wörtern gespeichert wird.

Die Daten sind vorzugsweise in einem nachhaltig standardisierten, menschenlesbaren Format gespeichert oder müssen vom Produkt ohne Verluste in ein solches geschrieben beziehungsweise von einem solchen gelesen werden können.

3.1.3 Produktleistungen

Das Produkt abstrahiert zwischen Daten (Model), Datenansicht (View) und Datenverarbeitung (Controlling). Es muss mit Datenmengen von wenigen bis ca. $4 \cdot 10^6$ Wörtern skalieren.

Die Enddaten sollen eine hohe Qualität aufweisen, so dass keine weitere Bearbeitung nötig ist. Trotzdem muss mit Fehlern gerechnet werden und es müssen entsprechende Routinen zur Behebung definiert sein (zum Beispiel Wiedereinspeisung eines fehlerhaft bearbeiteten Wortes).

Die Bereitstellung von Daten soll immer möglichst Augenblicklich geschehen. Sind aufwändigere Berechnungen nötig, so müssen diese so durchgeführt werden können, dass der restliche Betrieb nicht behindert wird.

3.2 Nicht funktionale Anforderungen

3.2.1 Benutzbarkeit

Das Produkt muss mit minimalem Aufwand auf allen gängigen Heimcomputer-Plattformen einsetzbar sein. Eine Internetverbindung wird vorausgesetzt. Es kann aber nicht vorausgesetzt werden, dass diese besonders schnell oder immer zur Verfügung steht.

3.2.2 Bedienbarkeit

Das Bedienkonzept muss so gestaltet sein, dass sich neue Benutzer schnell und einfach zurecht finden. Gleichzeitig sollen erfahrene Benutzer Zugriff auf besondere Funktionen haben, die die Arbeit erleichtern.

Die Bedienbarkeit soll im Sinne der Barrierefreiheit gestaltet sein, diese ist aber keine absolute Voraussetzung. Wenn möglich stützt sich die Bedienung auf bewährte und bekannte Konzepte.

3.2.3 Zuverlässigkeit

Normale Anforderungen

3.2.4 Effizienz

Normale Anforderungen

3.2.5 Änderbarkeit, Wartbarkeit

Das Produkt an sich beinhaltet keine Funktion, um sich selbst zu ändern. Änderungen werden von einem Systemadministrator/Systementwickler über ein Entwicklungswerkzeug vorgenommen.

Das Produkt beruht auf gängigen Standards und ist so kommentiert und dokumentiert, dass es auch von anderen Personen als deren Entwickler geändert oder gewartet werden kann.

3.2.6 Übertragbarkeit

Das Produkt muss so gestaltet sein, dass es auf verschiedene Plattformen übertragbar ist. Werden plattformspezifische oder proprietäre Funktionen verwendet, so sind diese so zu kapseln, dass sie leicht ersetzt werden können (Bsp.: Datenbankabstraktion).

4 Entwicklungszyklus

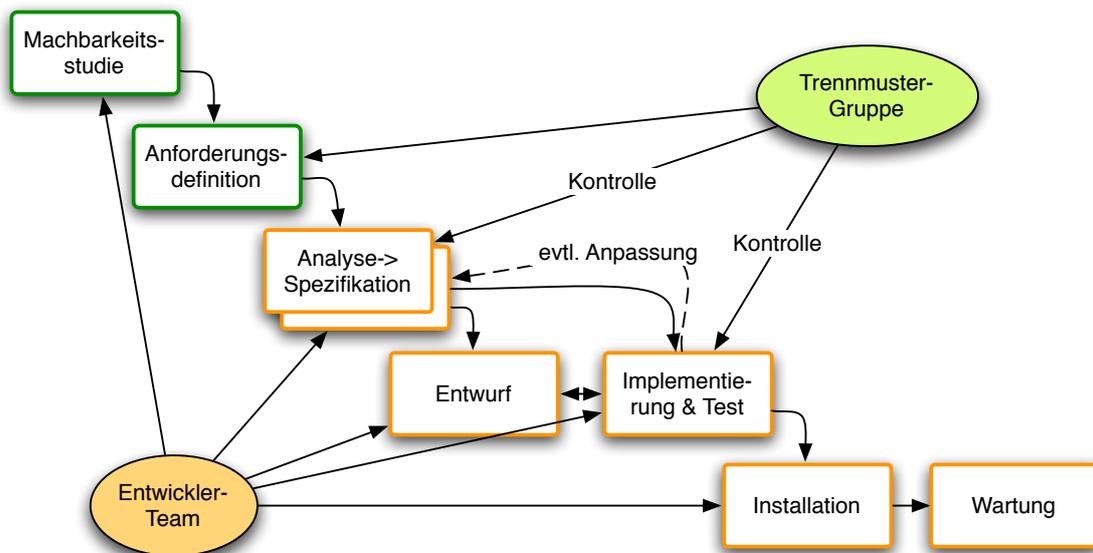


Abbildung 1: Vorgesehener Entwicklungszyklus. Die Gruppe hat in regelmässigen Abständen Einsicht in die Vorversionen. Falls nötig wird nach gegenseitiger Absprache die Spezifikation angepasst.

5 Use-Case

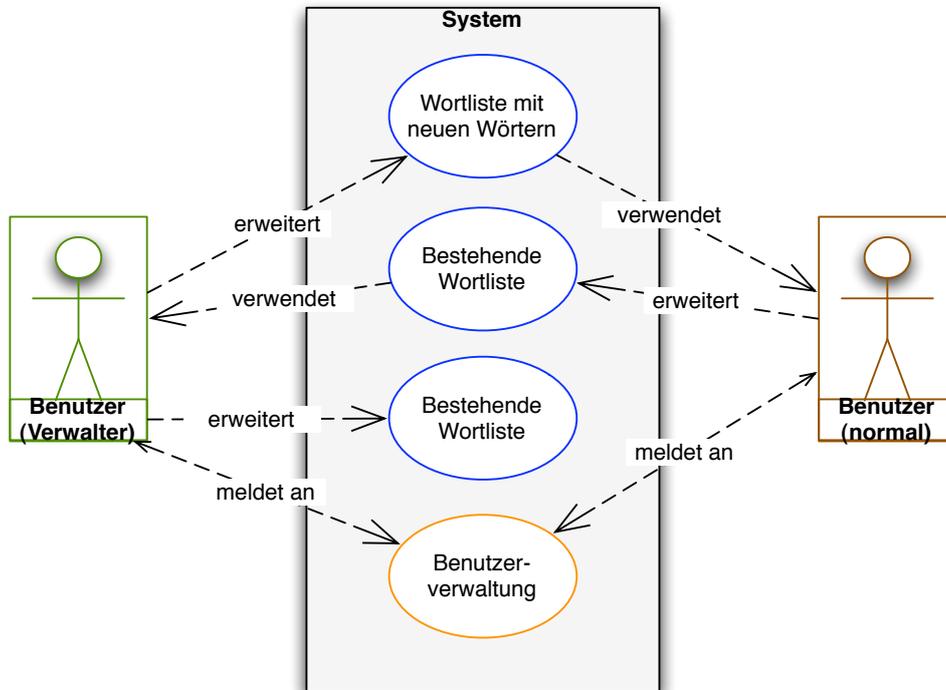


Abbildung 2: Anwendungsfälle des Produktes

6 Lieferumfang

TBD

7 Abnahmekriterien

TBD