

Verteilte Bearbeitung von Wortlisten.

Entwurf Version 2
Anforderungsspezifikation
(Lastenheft)

Mathias Nater

Mai 2009

trennmuster.org

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	5
3.1.3	Produktleistungen	5
3.2	Nicht funktionale Anforderungen	6
3.2.1	Benutzbarkeit	6
3.2.2	Bedienbarkeit	6
3.2.3	Zuverlässigkeit	6
3.2.4	Effizienz	6
3.2.5	Änderbarkeit, Wartbarkeit	6
3.2.6	Übertragbarkeit	7
4	Entwicklungszyklus	7
5	Use-Case	8
6	Lieferumfang	8
7	Abnahmekriterien	8

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.

Parallel dazu müssen die bestehenden Wortformen gepflegt werden.

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

33 Andererseits wird damit die bestehende Wortliste gepflegt.

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

36 Das Produkt wird auf sehr heterogenen Betriebsumgebungen verwendet.

37 **3 Produktfunktionen**

38 **3.1 Funktionale Anforderungen**

39 **3.1.1 Bereiche**

40 Das Produkt umfasst einen geschützten Verwaltungsbereich und einen öffentlichen Betriebsbe-
41 reich. Benutzer sind entweder dem Verwaltungs- und Betriebsbereich oder nur dem Betriebsbe-
42 reich zugeordnet.

43 Der Betriebsbereich ist in zwei klar getrennte Unterbereiche geteilt:

- 44 • ein Betriebsbereich „Hinzufügen“, wo neue Wortformen eingepflegt werden
- 45 • ein Betriebsbereich „Pflegen“, wo bestehende Wortformen und ihre Trennungen überarbei-
46 tet werden

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

48 **Im Verwaltungsbereich**

49 ... werden Benutzer verwaltet.

50 ... werden Wörter zur Bearbeitung eingegeben.

51 ... werden die Resultate der Bearbeitung abgeholt

52 ... werden Informationen zu laufenden Betrieb angezeigt

53 ... werden ausgewählte Einstellungen am Backend vorgenommen (z.B. häufige Fehltrennungen
54 erkennen, bearbeitete Wörter aus der Quellliste entfernen ...)

55 Im beiden Betriebsbereichen

56 ... werden bestimmte Einstellungen gemacht, die die Arbeit erleichtern (z.B. bisherige
57 Trennungen anzeigen, Wahl der Rechtschreibung, Eingabehilfen ein-/ausschalten)

58 ... werden bestimmte Informationen zum laufenden Betrieb angezeigt

59 ... können ganze Datenblöcke mit mehreren Wortformen bearbeitet werden

60 Im Betriebsbereich „Hinzufügen“

61 ... werden Wörter zur Bearbeitung ausgegeben

62 ... werden bearbeitete Wörter eingegeben

63 Im Betriebsbereich „Pflegen“

64 ... werden bestehende Wörter und ihre Trennungen korrigiert

65 3.1.2 Produktdaten

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

69 Die Daten sind vorzugsweise in einem nachhaltig standardisierten, menschenlesbaren Format
70 gespeichert oder müssen vom Produkt ohne Verluste in ein solches geschrieben beziehungsweise
71 von einem solchen gelesen werden können. Sie müssen in UTF-8 codiert sein.

72 3.1.3 Produktleistungen

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

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

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

82 **3.2 Nicht funktionale Anforderungen**

83 **3.2.1 Benutzbarkeit**

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

87 **3.2.2 Bedienbarkeit**

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

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

93 **3.2.3 Zuverlässigkeit**

94 Normale Anforderungen

95 **3.2.4 Effizienz**

96 Normale Anforderungen

97 **3.2.5 Änderbarkeit, Wartbarkeit**

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

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

102 Alle Dateien und sämtliche Daten sind in UTF-8 ohne ByteOrderMark (BOM) codiert, mit
103 UNIX-Zeilenwechsell (LF).

104 **3.2.6 Übertragbarkeit**

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

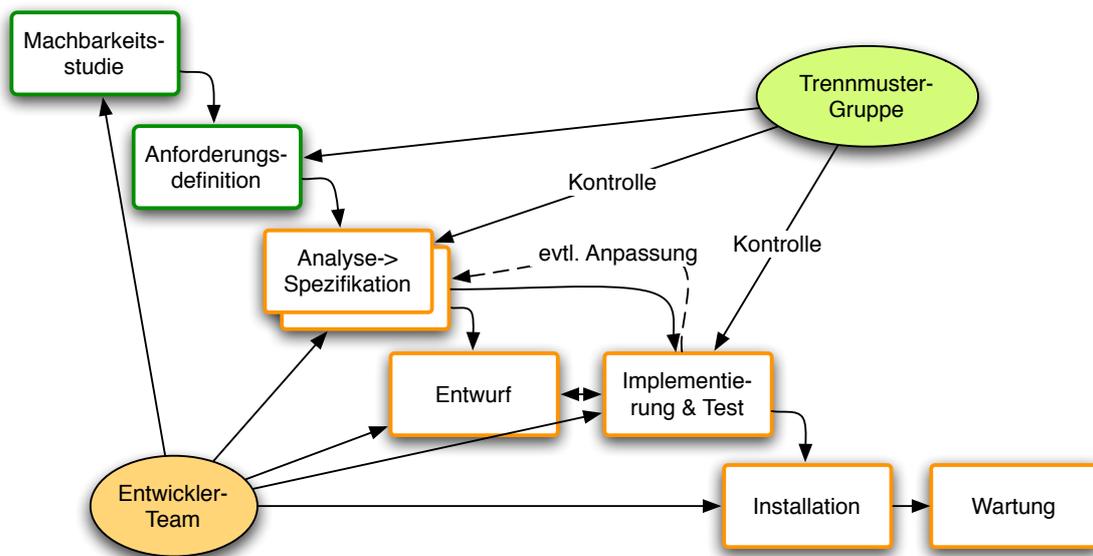
108 **4 Entwicklungszyklus**

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

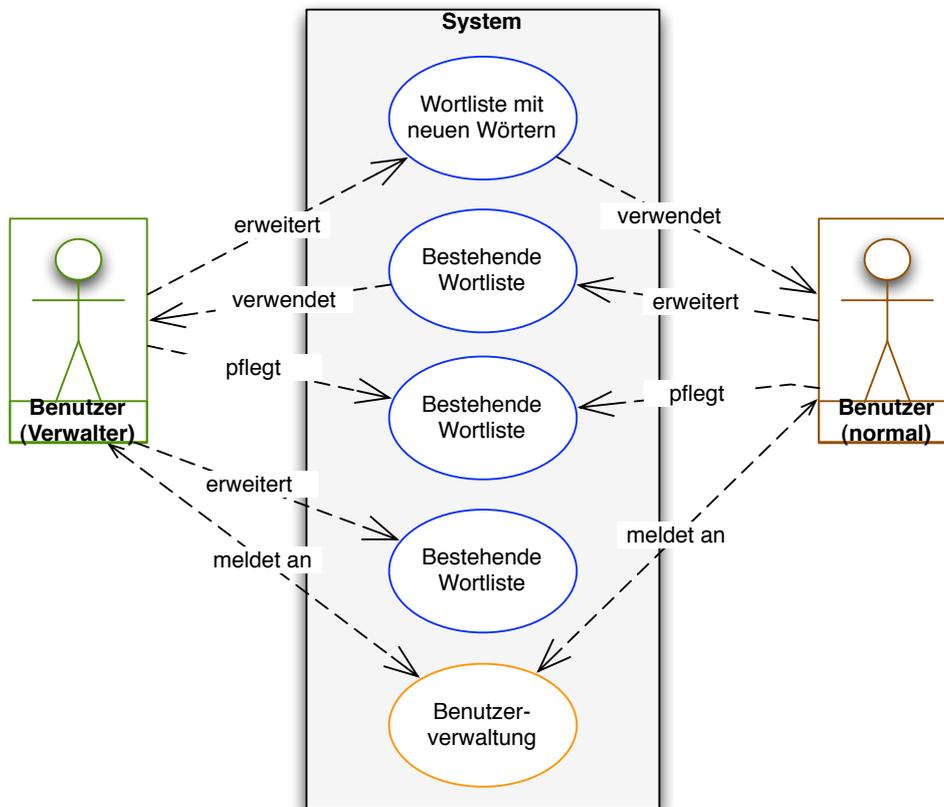
109 **5 Use-Case**

Abbildung 2: Anwendungsfälle des Produktes

110 **6 Lieferumfang**

111 TBD

112 **7 Abnahmekriterien**

113 TBD