

Institut für Informatik an der Universität Zürich  
Abteilung Computerlinguistik

Betreuung:  
Cerstin Mahlow

Dokumentation zum Programmierprojekt  
«Automatische Generierung von Kommentaren in SETEI»

Autor:  
David Lee  
Grünmattstr. 30  
8405 Winterthur  
052 232 77 51

## Inhalt

1. Idee und Zweck	3
2. Allgemeines zur technischen Realisation	3
3. Die inhaltlichen Kommentare	3
Bewertung	4
Korrektur	4
Ergänzung	4
Zurückspringen	5
Verweise	5
Modifikation der automatischen Kommentare	5
4. Die Verhaltenskommentare	6
5. Auswertung	7
Ergänzen der manuellen Kommentare	7
Verminderung der Kommentarzeilen	7
Fehlerhafte Kommentare	7
Angemessenheit der Bewertung	7
Qualitativer Vergleich mit manuellen Kommentaren	8
Fazit	8
6. Ausbaumöglichkeiten	8
7. Anhang	9
Algorithmus zur Erzeugung der inhaltlichen Kommentare	9
Manuelle Kommentare aus der Übung «Lemma»	10
Automatische Kommentare aus der Übung «Lemma»	12

## 1. Idee und Zweck

Ziel des Programmierprojektes war es, die Java-Applikation SETEI so zu erweitern, dass Kommentare automatisch erzeugt werden können. Ausgangslage für die Idee war die Beobachtung, dass viele in SETEI von Hand geschriebene Kommentare einem stereotypen Muster folgten: Zuerst einmal eine allgemeine Bewertung wie «Ausgezeichnet», «Falsch», «Richtig» etc., gefolgt von einer Richtigstellung der falschen Aussagen und einer Ergänzung von fehlenden Aussagen. Der meist gleiche Aufbau eines Kommentars sollte eine Automatisierung einfach machen: Vom Benutzer vergebene Punktzahlen lassen sich zusammenzählen und in eine Bewertung in Worten übersetzen. Der Satz mit der höchsten Punktzahl dient dann als Referenz für Richtigstellungen im Kommentar.

Von automatischen Kommentaren waren folgende Vorteile zu erhoffen:

- Abnehmen der Grundarbeit: der Benutzer braucht die Kommentare nur noch anzupassen.
- Garantie, dass jeder Satz kommentiert ist.
- Bei automatisch erstellten Kommentaren werden natürlich auch deren Bedingungen automatisch generiert. Somit wird der Benutzer vom mühsamen Zuordnen der richtigen Zeilennummern und logischen Operatoren befreit.
- Dadurch, dass die Kommentare jederzeit gelöscht und neu generiert werden können, werden in den Bedingungen tote Referenzen verhindert.

Neben den für jede Übung spezifischen Kommentaren können seit kurzem noch andere erzeugt werden, die unabhängig von der Übung immer gleich aussehen. Es handelt sich um Kommentare, die aufgrund des Benutzerverhaltens beim Durchführen der Übung ausgegeben werden. Das heisst, es wird nicht eine bestimmte Kombination von Elementen kommentiert (z.B. 1 & 4 & 16), sondern zum Beispiel geprüft, wie viel Mal einzelne Elemente vorkommen oder wie lange der Gesamtpfad ist. Durch zirkuläre Verbindungen und Backtracking können Elemente während einer Übung mehrmals ausgewählt werden. Für solche Kommentare sollte im Programmierprojekt ebenfalls ein Automatismus entwickelt werden.

Ich nenne diese Art von Kommentare im Folgenden «Verhaltenskommentare», die anderen hingegen «inhaltliche Kommentare».

## 2. Allgemeines zur technischen Realisation

Elegant wäre es gewesen, die ganze Erweiterung in separaten Dateien/Klassen zu programmieren, so dass diese Klassen SETEI einfach hätten hinzugefügt werden können. Es wäre auch für die Dokumentation einfacher gewesen, das ganze Programmierprojekt als separates Stück Code zu haben.

Leider war das nicht möglich. Es mussten diverse Anpassungen in bestehenden Klassen gemacht werden:

- Die Klasse «MenuManager», welche die Menüleiste erstellt, musste um das Menü «AutoKommentar» erweitert werden.
- Die Datei «menus.set», welche die Menübefehle speichert, musste natürlich ebenfalls angepasst werden.
- Die Klasse «EdgeData», welche ein Element/eine Zeile in einer Box repräsentiert, musste um das Attribut «autoGenerated» erweitert werden, um zu überprüfen, ob ein Kommentar automatisch erzeugt wurde oder nicht.
- Die grosse Methode «Commands.allPhrases()» musste grundlegend geändert werden. Die Methode liefert alle Sätze zurück, die mit dem gegenwärtigen Automaten gewählt werden können. Bisher geschah das einfach in Form einer Zeichenkette, die dann in einem Fenster gezeigt und ggf. abgespeichert werden konnte. In dieser Form können die Sätze aber nicht analysiert werden. Daher gibt es jetzt eine eigene Klasse namens «Phrase», welche einen Satz repräsentiert. Diese Klasse enthält alle Kanten des Satzes sowie, getrennt davon, alle Kanten des Kommentars. Darüber hinaus enthält die Klasse eine Anzahl Methoden, die wichtige Informationen eruieren, z.B. die Gesamtpunktzahl, den Pfad (ID-Nummern der Satzteile), alle falschen Satzteile, alle Satzteile mit 0 Punkten, etc. `Commands.allPhrases()` gibt nun einen Vector aus Phrases-Objekten zurück.
- Die Kommentare für das Benutzerverhalten werden in einer neuen Art von Box gezeigt. Dafür habe ich eine Unterklasse von «CommentBox» kreiert und «BehaviourCommentBox» genannt. Die Methoden zum Laden und Speichern mussten auf diese neue Box-Klasse und die veränderte EdgeData-Klasse eingestellt werden, und da Verbindungen zu dieser Box unsichtbar sind und automatisch geknüpft werden, musste auch die Klasse «Connections» angepasst werden.

Wenigstens wurden die Hauptroutrinen für die Erzeugung der Kommentare in einer separaten Klasse namens «CommentGenerator» platziert.

Über bestehende SETEI-Komponenten gibt die SETEI-Dokumentation Auskunft.

## 3. Die inhaltlichen Kommentare

Die auf den Inhalt der Übung Bezug nehmenden Kommentare werden mit dem Menübefehl «Kommentare erzeugen» generiert. Die dafür zuständige Methode ist `generateComments()` in der Klasse `CommentGenerator`. Sie generiert zuerst alle Sätze mit den entsprechenden Punktzahlen. Dann erzeugt sie für jeden Satz einen String mit dem Kommentartext. Bewertung, Korrektur und Linkverweise werden fortlaufend am selben String angehängt. Wenn für jeden Satz ein

String erzeugt worden ist, wird für jeden String eine Kommentarzeile geschaffen und in einer (vorhandenen oder wenn nötig neu generierten) Kommentarbox eingefügt.

Bei einer Neuerzeugung löscht das Programm allfällige alte automatisch erstellte Kommentare. Ein automatisch erzeugter Kommentar verliert das Attribut «autoGenerated», wenn er nachträglich vom Benutzer geändert wurde. Das bedeutet, dass automatisch erzeugte, aber von Hand geänderte Kommentare nicht überschrieben werden.

Die Erzeugung des Kommentartextes gliedert sich in vier Teile: Bewertung, Korrektur, Ergänzung und Linkverweise. Eine Bewertung erfolgt immer, eine Korrektur immer ausser beim besten Satz, Ergänzung und Linkverweise nur in gewissen Fällen.

### Bewertung

Die Bewertung erfolgt aufgrund der totalen Punktzahl des gewählten Satzes sowie aufgrund dessen, ob der Satz Elemente mit negativer Punktzahl (=falsche Elemente) enthält oder nicht. Ist ein solches Element vorhanden, lautet die Bewertung «Nicht ganz», sind es mehrere, lautet sie «Falsch». Sind keine falschen Elemente vorhanden, wird «Richtig» ausgegeben. Sind weder falsche noch richtige Elemente drin, sondern ausschliesslich solche mit 0 Punkten, so lautet die Bewertung «Nicht falsch, aber nichtssagend.» Der beste Satz erhält «Ausgezeichnet», der schlechteste «Unsinn».

### Korrektur

Bei der Korrektur wird für jede Kante im Satz ermittelt, ob es eine bessere Wahl gegeben hätte. Wenn ja, wird in einem neuen Satz die falsche Aussage durch diejenige mit der höchsten Punktzahl ersetzt und der ersetzte Teil hervorgehoben. Wenn nein, wird einfach angemerkt, dass diese Wahl falsch ist, sofern die Punktzahl negativ ist.

Die Bewertung und die Korrektur werden sprachlich so verknüpft:

- Richtig, aber dies wäre noch besser gewesen: (+ korrigierter Satz)
- Falsch. (+ korrigierter Satz)

Eine Kante kann aber nicht immer durch eine beliebige andere ersetzt werden, da in den Zeilen der Stepboxen Bedingungen stehen können. In der Übung «Lemma» ist das zum Beispiel der Fall.

12: Ein Lemma ist das, was nach dem Prozess | 17: der Stemmtisierung (...) | übrig bleibt.

Anstelle von Zeile 12, die drei Punkte gibt, kann man die Zeile 1 wählen, welche 5 Punkte gibt. Dann würde der Satz aber lauten:

1: Ein Lemma ist | 17: der Stemmtisierung (...) | übrig bleibt.

Das ist unsinnig und von den Bedingungen ausgeschlossen. Daher wird vor dem Ersetzen überprüft, ob die nachfolgenden Zeilen die zu ersetzende erlauben. Wenn nicht, wird (ggf. wiederholt) zu der nächstbeseren Zeile gesprungen.

Auf diese Weise wird zwar nicht der bestmögliche Satz ausgegeben (respektive nur dann, wenn die Suche nach einer geeigneten Kante in keinem Satzabschnitt einen Erfolg hat), aber das wird im Kommentar auch nicht behauptet. Es ist der beste Satz *innerhalb derjenigen Möglichkeiten, die der Student auch zur Auswahl bekommen hat*. Es bringt ja wenig, einen Satz als Korrektur vorzusetzen, dessen Elemente der Student gar nie gesehen hat, weil er einen anderen Pfad durch die Boxverbindungen gegangen ist.

### Ergänzung

Nun wird die Länge des bestmöglichen Satzes mit der Länge des gerade behandelten Satzes verglichen. Ist der bestmögliche länger, so hat der Student einige Dinge ausgelassen, und es erfolgt die Bemerkung «Man hätte noch erwähnen können:» plus der bestmögliche Satz mit den fehlenden Teilen hervorgehoben. Ist der beste Satz hingegen kürzer, so lautet der Kommentar «Man hätte das kürzer und präziser formulieren können:», worauf ebenfalls der beste Satz folgt.

Man mag kritisieren, dass es schwerfällig sei, immer den ganzen Satz zu wiederholen statt nur die fehlerhaften Teile. Es ist jedoch nichts anderes möglich, wenn man keine Syntaxanalyse für den Satz vornehmen kann. Am besten lässt sich das anhand eines Beispiels verdeutlichen.

Angenommen, der Automat sieht wie auf der Abbildung 1 aus: und der Schüler wählt nun «Grimmelshausen war ein Schriftsteller im Mittelalter.», so lautet der erzeugte Kommentar:

«Nicht ganz. Grimmelshausen war ein Schriftsteller **im 17. Jahrhundert.**»



Abbildung 1

Wenn nun nur die falschen Teile ausgegeben würden, so wäre der Kommentar: «Nicht ganz. im 17. Jahrhundert», oder wenn die Startbox noch miteinbezogen wird: «Nicht ganz. Grimmelshausen war ein im Mittelalter.»

Wäre der Kommentar von Hand geschrieben worden, würde er wohl lauten «Nicht ganz. Grimmelshausen lebte im 17. Jahrhundert.» Hier kommt sogar ein ganz neues Wort hinzu («leben»). Solche Kontextanpassungen sprengen den Rahmen eines einfachen Programms, das lediglich Zeichenketten zusammensetzt, bei Weitem.

Die Boxenzugehörigkeit gibt keinen Aufschluss über die Zugehörigkeit eines Elements zu einer bestimmten grammatikalischen Konstituente. Es ist zum Beispiel nicht zwingend, dass die ersten beiden Boxen einen vollständigen Satz bilden. Die Auswahlmöglichkeiten der Boxen müssen nicht einmal der selben syntaktischen Kategorie angehören. Ohne Syntexanalyse ist es darum nicht möglich, Teile wegzulassen. Es ist auch nicht möglich, Konstruktionen mit anderer Wortstellung zu erzeugen wie z.B. Nebensätze: «Man hätte noch erwähnen können, dass...»

Korrektur und Ergänzung heben die fehlerhaften Teile mit dem HTML-Tag <B> hervor. Die Ausgabe muss deshalb in HTML-Umgebung erfolgen, und die Kommentare dürfen nicht generell fett gedruckt sein, da man sonst die Hervorhebung nicht sieht.

### Zurückspringen

Wenn der Satz über einen Pfad ging, der zu kurz war, d.h. wenn wichtige Dinge verpasst wurden, so kann optional der Student zum Verzweigungspunkt zurückgeschickt werden anstelle einer Ergänzung im Kommentar. Diese Wahl zwischen den beiden Modi wird zu Beginn der Kommentarerzeugung gemacht. Technisch funktioniert das ganz einfach, indem in der jeweiligen Kommentarzeile eine Followup-Box angegeben wird.

### Verweise

Nach der Korrektur kann der Kommentar noch um einen Hinweis auf einen oder mehrere andere Satztests erweitert werden. Der Hinweis wird dann eingefügt, wenn im ausgewählten Satz ein Wort einer Liste mit Schlüsselwörtern vorkommt und lautet zum Beispiel so:

Es gibt auch eine Übung zum Thema [natürlichsprachliche Beispiele für reguläre Sprachen](#).

Beim unterstrichenen Wort handelt sich um einen HTML-Link, der zum Start der entsprechenden Übung führt.

Das Verfahren ist in diesem Programmierprojekt sehr primitiv und in der Praxis wohl kaum zu gebrauchen. In einer Textdatei namens «keywords.txt» befindet sich eine Liste mit allen Übungen, angeordnet im Format *Perl-Programmname:Suchbegriff1\*Suchbegriff2:Beschreibung*

als Beispiel:

bspregsprache:reguläre Sprachen\*reguläre Sprache\*regulärer Sprache:natürlichsprachliche Beispiele für reguläre Sprachen

Findet SETEI innerhalb des vom Studenten gewählten Satzes einen der Suchbegriffe, so wird ein Verweis angefügt, bei dem die Beschreibung der sichtbare Text des Links ist und der Programmname eine Perl-Angabe im HTML-Link.

Damit die entsprechenden Stellen sicher gefunden werden, müssen bei den Suchbegriffen die Deklinationsformen einzeln aufgeführt sein. In SETEI wird zudem eine Normalisierung vorgenommen: alle Buchstaben werden klein gesetzt und alle Umlaute zu ae, ue und oe konvertiert. Mit der Umsetzung in Kleinbuchstaben sind allerdings Abkürzungen in den Suchbegriffen noch problematischer, als sie ohnehin schon sind. Ich hatte am Anfang zum Suchbegriff «Information Retrieval» noch «IR» hinzugefügt, aber die Buchstabenkette «ir» taucht des öfteren in Texten auf und somit auch der völlig deplatzierte Verweis auf die Übung zu Information Retrieval.

Bei den mitgelieferten Testübungen wird jedesmal ein solcher Hinweis erzeugt, weil diese Übungen bereits in der Liste aufgeführt sind.

### Modifikation der automatischen Kommentare

Unter Umständen lässt ein Satztest eine sehr grosse Anzahl möglicher Sätze zu, und dann werden auch entsprechend viele Kommentare generiert. Das ist zumindest unschön und beeinträchtigt die Geschwindigkeit der Anwendung. Auch das nachträgliche Bearbeiten ist nicht ideal: Zwar kann man sehr schön einen Satz herausuchen und genau diesen spezifischen Satz bearbeiten, aber eine allgemeine Korrektur (z. B. Ändere generell «Richtig» in «Gut» ab), ist nicht so einfach möglich. Es gibt zwar in SETEI eine Suchen/Ersetzen-Funktion, aber dabei besteht die Gefahr, dass auch Textstellen ersetzt werden, die man nicht ersetzen wollte.

Ich habe die Methode «ein Kommentar pro Satz» gewählt, weil sie am nahe liegendsten war – ich sehe auch jetzt keine andere Möglichkeit als die, jeden möglichen Satz durchzugehen. Ausserdem bringt diese Methode den Vorteil, dass sicher jeder Satz kommentiert wird. Die so erzeugten Kommentare nachträglich effizienter anzuordnen, ist kein Problem. Ich habe den ersten Schritt in diese Richtung bereits in diesem Programmierprojekt gemacht mit der Funktion «Kommentare zusammenfügen». Ruft man diesen Menübefehl auf, so werden alle Kommentare, die den selben Text haben, zu einem Kommentar zusammengefasst. Die Bedingungen werden erweitert nach dem Muster:

(Bedingung erster Satz) ODER (Bedingung nächster Satz) ...

Hat zum Beispiel der eine Kommentar die Bedingung 1&5&8 und der andere die Bedingung 1&4&9, so erhält der neue Satz die Bedingung (1&5&8)#(1&4&9).

Eine weiter gehende Umformung bestünde darin, dass für die drei Bestandteile eines Kommentars jeweils separate, mit einander verknüpfte Boxen erzeugt werden. Das war ursprünglich auch so geplant. Es gäbe dann eine Box mit allen Bewertungen, eine mit allen Korrekturen und eine mit allen Verweisen. Dadurch würde allerdings die Anzahl der Kommentare nur unwesentlich geringer als mit der bereits realisierten Lösung, denn die Elemente der Boxen lassen sich nicht einfach beliebig kombinieren. So folgt auf die Bewertung «Ausgezeichnet» überhaupt keine Korrektur, und auf die Bewertungen «Nicht ganz», »Richtig, aber» und «Falsch» erfolgen jeweils andere Korrekturen.

Gegen das Konzept mit mehreren Boxen spricht noch etwas anderes. Wenn der Benutzer einen Kommentar ändert, wird dieser als nicht mehr automatisch erzeugt angesehen, d. h. er wird beim Neugenerieren nicht überschrieben. Was soll geschehen, wenn der Benutzer nun nur die Korrektur ändert? Oder nur die Bewertung? Die Anpassung wäre zumindest sehr kompliziert. Auch die Kommentare, die der Benutzer erstellt hat, müssten in dieses komplizierte System (wie das auch immer aussehen soll) einbezogen werden.

Mit dem Zusammenfügen sowie einer anderen Erweiterung ist SETEI nun in der Lage, sowohl Kommentare für einen einzelnen Satz zu ändern als auch gleich lautende Kommentare für mehrere Sätze. Um automatische Kommentare nicht an einem spezifischen Ort, sondern ganz generell zu ändern, können alle Texte, die der Kommentargenerator verwendet, in einem Dialogfenster editiert werden. Dieser Dialog, der mit dem Menü «Textbausteine...» aufgerufen wird, erlaubt auch den Import und Export dieser Texte. So kann man beispielsweise die Kommentare für andere Sprachen als Deutsch verwenden, oder den Studenten generell mit «Sie» statt mit «Du» anreden.

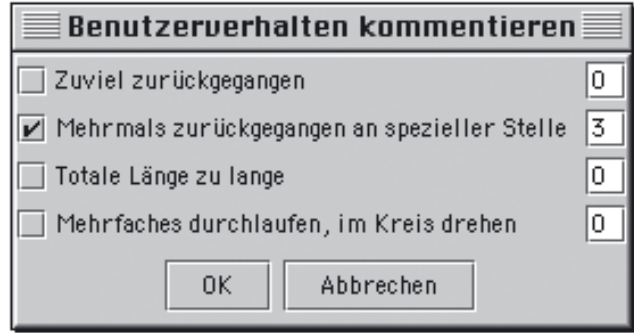


Abbildung 3

Wie die Abbildung 2 zeigt, lassen sich solche Sets von Textbausteinen auch abspeichern und wieder importieren.

#### 4. Die Verhaltenskommentare

Diese Kommentare werden über den Menüpunkt «Benutzerverhalten kommentieren» editiert. Der Dialog in Abbildung 3 erscheint, bei dem die verschiedenen Möglichkeiten aktiviert oder deaktiviert werden können. Der jeweilige Schwellenwert wird hier angegeben. Bestätigt man seine Eingaben, so erscheinen in einer speziellen Box, der «BehaviourCommentBox» die entsprechenden Kommentare. Im gezeigten Beispiel wird der Kommentar «An einer Stelle hast du wohl einfach geraten» ausgegeben, wenn an einer Stelle mehr als drei mal zurückgesprungen wurde. Der ausgegebene Text kann unter dem Menüpunkt «Textbausteine» (Abbildung 2, unterste vier Zeilen) definiert werden. Die Bedingungen definieren diese Art von Kommentaren als solche, indem sie das Prologprädikat «ho» (für «higher order») verwenden.

Es spricht nichts dagegen, diese Kommentare wie normale Kommentarzeilen zu editieren. Der Dialog ist

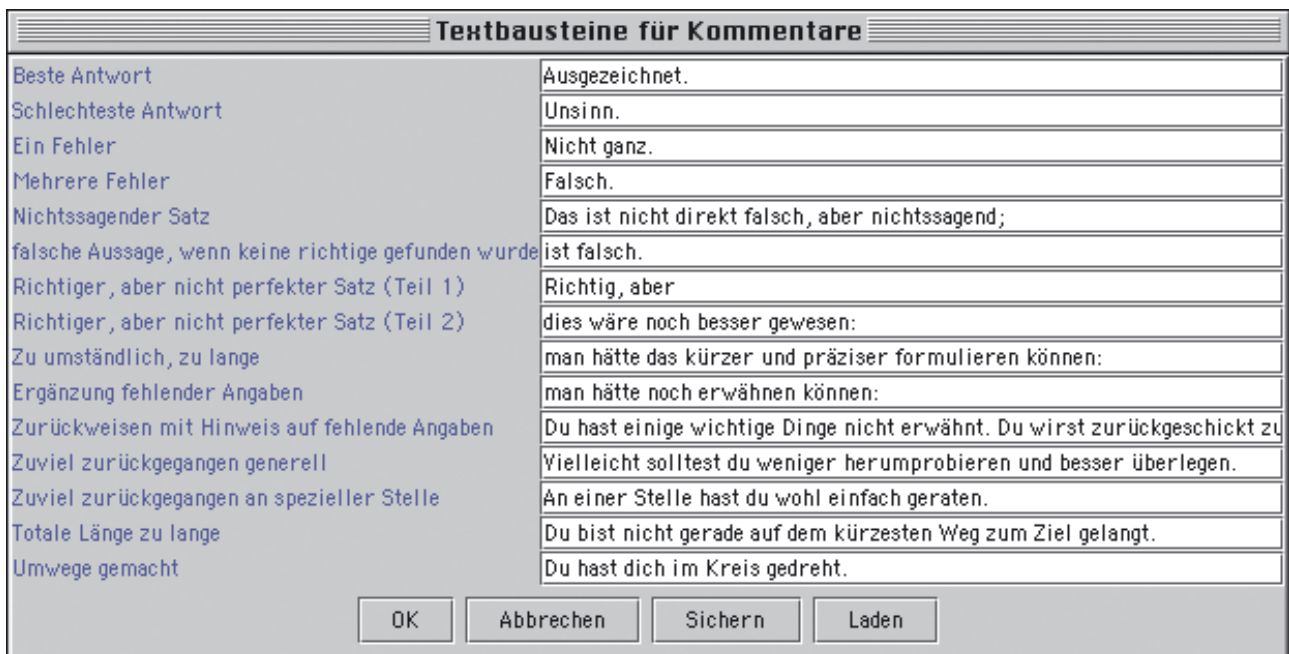


Abbildung 2

nur eine zusätzliche Hilfe.

Die BehaviourCommentBox ist automatisch mit jeder Endbox verknüpft, welche keine Kommentarbox ist. Da diese Verknüpfungen zahlreich sein können und da der Benutzer sie nicht ändern soll, sind diese Verknüpfungen unsichtbar. Im Boxbrowser sieht man aber, welche Boxen zu ihr hinführen.

Man kann die Eigenschaft der automatischen Verknüpfung dazu «missbrauchen», weitere, gewöhnliche Kommentare von Hand in die BehaviourCommentBox zu schreiben, wenn man sichergehen will, dass diese Kommentare garantiert nach jedem Satz erscheinen sollen.

## 5. Auswertung

Zum Zeitpunkt der Entwicklung des Kommentargenerators waren noch keine Satztests erstellt worden, die Punktzahlen enthielten, da es sich um Lernübungen und nicht um Prüfungsaufgaben handelte. Daher hat Esther Kaufmann, welche die Übungen zu den Einführungsvorlesungen der Computerlinguistik entwirft, drei Satzgeflechte im Nachhinein mit Punktzahlen versehen.

### Ergänzen der manuellen Kommentare

Eine mögliche Anwendung des Kommentargenerators wäre es ja, die vom Autor vergessenen Sätze mit einem Standardkommentar zu versehen. Bei den kleineren, übersichtlichen Übungen «Lemma» und «Morphologie» war das nicht nötig, alle Sätze (36 bzw. 46 an der Zahl) wurden von Hand kommentiert.

Die Übung «Parser» ist komplexer aufgebaut und lässt 168 mögliche Sätze zu. Die Autorin hat für diese Übung einen Standardkommentar von Hand geschrieben, dessen Bedingungen so angelegt sind, dass dieser Kommentar immer dann erscheint, wenn kein anderer Kommentar gewählt wird. Dadurch ist garantiert, dass jeder Satz kommentiert ist. Ohne diesen Trick würde rund die Hälfte der Sätze nicht kommentiert, und der Kommentargenerator könnte einspringen.

Da im Moment die Übungen weiterhin ohne Punktzahlen gemacht werden und ein solcher manueller Standardkommentar recht lange Bedingungen aufweist, könnte in SETEI ein Befehl eingebaut werden, der dessen Bedingungen automatisch einsetzt.

### Verminderung der Kommentarzeilen

Von den zahlreichen automatisch erstellten Kommentaren kann eine gewisse Zahl durch das Zusammenfügen gleich lautender Kommentare eliminiert werden. Das reduziert die Anzahl allerdings nur auf die Hälfte bis einen Drittel der ursprünglichen Kommentarzeilen. Zum grössten Teil lässt sich dies damit erklären, dass es viele Kommentare gibt, die zwar den genau gleichen Wortlaut haben, aber nicht die gleiche Stelle fett markiert (es gibt sehr viele mögliche Kombinationen, was man alles falsch auswählen kann). Daher sind die vielen Sätze nicht einfach als überflüssigen

Ballast anzusehen, denn im Gegensatz zu den von Hand erstellten Kommentaren zeigen die automatischen dem Studenten viel genauer an, was er falsch ausgewählt hat. Die von Hand erstellten müssen sich auf die grössten Fehler des Satzes konzentrieren, da man von Hand nicht die Übersicht und die Zeit hat, mehrere hundert Möglichkeiten zu kommentieren.

### Fehlerhafte Kommentare

Bei Automaten, in denen «Abkürzungen» wie in Abbildung 4 genommen werden können, erzeugt der Kommentargenerator für gewisse Sätze Kommentare doppelt. Der Grund: Auf Abbildung 4 gelangt man zur Box 4 sowohl direkt von Box 2 als auch auf Umwegen über Box 3 oder 6. Dadurch entsteht zum Beispiel ein Satz mit den Elementen 1,2,7 und 20, dessen Bedingungen (1&2&7&20) aber auch im Satz 1/2/3/7/20 (und weiteren) erfüllt sind. Die Behebung dieses «Bugs» ist nicht ganz trivial. Die Bedingung (1&2&7&20) müsste erweitert werden zu 1&2&7&20&~(3#10#19#20). Alle Zeilen, die in den «zusätzlichen» Boxen sind, sind damit negiert. Das setzt aber voraus, dass die Methode zur Kommentargenerierung auf ein Modell des Boxengeflechts zugriff hat, was momentan nicht der Fall ist. Es gibt im Moment in SETEI gar keine interne Repräsentation der «Boxenbäume».

In den Testsätzen kam ein solcher Automat nicht vor. «Parser» hat zwar diese Baumstruktur, doch die hier erwähnten Sätze können aufgrund der Bedingungen gar nicht gewählt werden.

### Angemessenheit der Bewertung

Beim ersten Testlauf mit der Übung «Lemma» erstellte das Programm ausschliesslich Kommentare, die «richtig» oder «nicht ganz» als Bewertung enthielten, während die entsprechenden von Hand geschriebenen Kommentaren viele dieser Sätze als falsch taxierten. Ähnliches lässt sich auch über die beiden anderen Satztests sagen. Der Grund dafür liegt darin, dass falsche Selektionen nicht immer mit negativen Punktzahlen versehen sind. Der Satz

|0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |34: von einem Wort übrig bleibt.

Wird kommentiert mit:

Nein. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Es kann allerdings der Fall eintreten, dass nach einer Trunkierung zufälligerweise genau das Lemma einer Wortform übrig bleibt.

Da ein Lemma nichts mit Trunkierung zu tun hat, müsste Zeile 13 («der Trunkierung») mit einem negativen Wert belastet sein. Sie hat aber eine positive Punktzahl. Dasselbe gilt für die beiden Zeilen aus Box 2. Sie haben zwar nur einen Punkt, im Gegensatz zu den richtigen, denen fünf Punkte gegeben wurde,

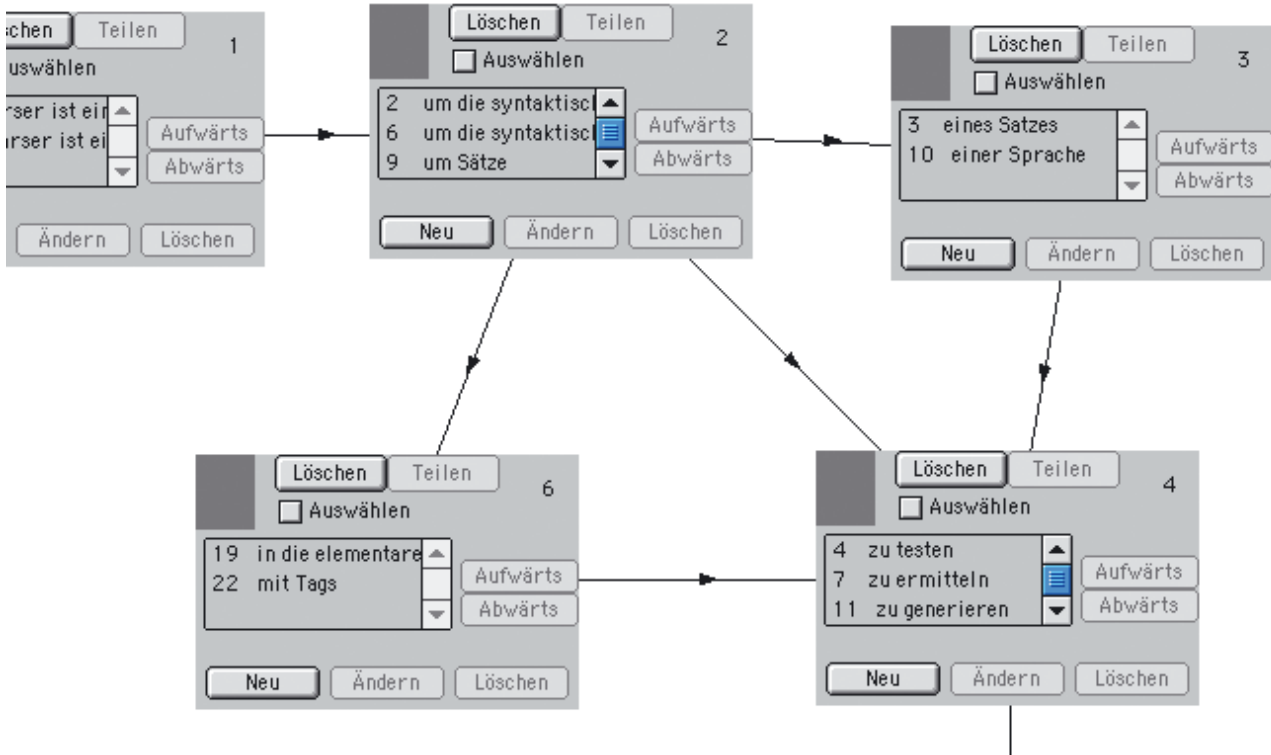


Abbildung 4

aber solange die Zahl nicht negativ ist, gilt die Wahl als richtig.

Nach der Korrektur der Punktzahlen erfolgte die automatische Bewertung einigermaßen übereinstimmend mit der manuellen. Es muss also dem Anwender (z.B. in der Hilfedatei) deutlich gesagt werden, wie die Punktzahlen vergeben werden sollen.

### Qualitativer Vergleich mit manuellen Kommentaren

Wenn man die im Anhang aufgeführten manuell erstellten Kommentare der Übungen betrachtet, sieht man sehr schön, dass der Aufbau dem automatischen Schema meist sehr ähnlich ist.

Ein Kommentar beginnt zuerst mit einer Bewertung, gefolgt von der Korrektur in Form der richtigen Definition. Manchmal sind die Bewertungen und Korrekturen sprachlich angepasst («Das ist ein Parser sicher nicht», «Denn ein Parser ist *keine statische Beschreibung, sondern eine Prozedur, [...]*»). Zur Abwechslung sind auch Variationen eingebaut («Das ist total falsch» vs. «Das stimmt ganz und gar nicht»). Solche Dinge kann der Kommentargenerator nicht leisten, aber inhaltlich sollten die Kommentare etwa dasselbe aussagen.

Einige Kommentare enthalten noch eine Zusatzinformation, quasi als Weiterbildung: *Die Bezeichnung «Parser» kommt vom lateinischen Wort «pars», was «Teil» bedeutet.* Dies ist ein Typ von Kommentar, welcher sowieso nur von Hand erzeugt werden kann.

Ob die Qualität für den Einsatz in der Praxis genügt, kann ich hier nicht beurteilen. Natürlich müssen die automatischen Kommentare dazu nicht den manuellen ebenbürtig sein, denn sie sollen diese nicht ersetzen, sondern lediglich die manuellen ergänzen oder als Grundstein für manuelle Einträge dienen.

### Fazit

Das Vorhaben des Projekts konnte umgesetzt werden, die Kommentare werden in der Art und Weise erzeugt, wie das geplant war. Einziger bekannter Fehler ist das bereits erwähnte Erzeugen der doppelten Kommentare.

## 6. Ausbaumöglichkeiten

In SETEI sollte (nicht nur wegen des Kommentargenerators) eine Klasse erstellt werden, welche die Boxenwege repräsentiert. Dadurch könnte, wie oben gezeigt, der Fehler mit den doppelten Sätzen behoben werden.

Für die Stichwortliste könnte man sich überlegen, Gertwol einzubinden.

Damit nicht alle Kommentare den gleichen Wortlaut haben, könnte man sprachliche Variationen mit der selben Bedeutung hinzufügen, beispielsweise («Richtig», «Korrekt», «Ja», «Gut»).

Schliesslich könnte man Kommentare auch grundsätzlich anders generieren als durch Zusammensetzen von fertigen Strings. Um grammatikalisch angepasste Kommentare (z.B.: «Es ist richtig, dass ...») zu erzeugen, müsste jedoch zumindest ein Tagger, besser ein Parser in SETEI eingebunden werden.



## Anhang

### Algorithmus zur Erzeugung der inhaltlichen Kommentare

#### 1. Vorbereitung

Erzeuge alle möglichen Sätze

Gehe alle Sätze durch und bestimme diejenigen mit der höchsten und niedrigsten Punktzahl.

Wenn kein Satz mehr als 0 Punkte hat, Warnung anzeigen und abbrechen.

Dialogfenster für Optionen (Fall1 oder Fall2 unter 2.3) öffnen.

Wenn nicht vom Benutzer abgebrochen wurde, alte Kommentare aus den Boxen löschen.

#### 2. Erzeugung des Kommentartextes

##### 2.1. Gesamtbewertung

wenn Punktzahl = Maximum : Text für besten Satz. Keine Korrekturen/Ergänzungen.

sonst:

wenn keine falschen Teile:

wenn Punktzahl grösser als 0 : Text für richtige Aussagen.

wenn Punktzahl gleich 0: Text für nichtssagende Aussagen.

sonst:

wenn nur ein falscher Teil: Text für nicht ganz richtige Aussagen.

wenn mehrere falsche Teile: Text für falsche Aussagen.

##### 2.2. Korrektur falscher und nicht optimaler Aussagen

Erzeuge einen neuen Satz newPhrase, welcher eine Kopie des gegenwärtigen ist

für jede Kante im Satz

ermittle die Kante in der gleichen Box mit der höchsten Punktzahl

wenn maximale Punktzahl grösser als 0 und die Kante nicht gleich der gegenwärtigen:

ersetze in newPhrase die falsche Kante durch die beste und markiere diese fett.

sonst: bezeichne die gewählte als falsch.

füge dem Kommentartext den neuen Satz an, sofern mindestens eine Kante ersetzt werden konnte

sonst füge den besten Satz als Korrektur an

##### 2.3. Ergänzung fehlender bzw. kürzer formulierter Aussagen

wenn der Satz kürzer ist als der beste

Fall1: der Benutzer will Korrektur im Kommentar

häng die entsprechende Bemerkung an den Kommentartext

häng den besten Satz an, wobei die nicht erwähnten Teile mit positiver Punktzahl fett hervorgehoben werden.

Fall2: der Benutzer will, dass der Student den fehlenden Teil wiederholen muss

häng Bemerkung an den Kommentartext, dass zurückgesprungen wird

vergleiche die Box-Nummern des besten Satzes mit denjenigen dieses Satzes

speichere die erste Nummer des besten Satzes, die nicht übereinstimmt, als Followupbox.

wenn der Satz länger ist als der beste

häng die entsprechende Bemerkung an den Kommentartext

häng den besten Satz an

speichere den Kommentartext und die Followup-Box.

#### 3. Verweise im Kommentar erzeugen

wandle den Text der Zeile in Kleinbuchstaben um

ersetze alle ä/ö/ü durch ae/oe/ue

lies die Liste mit den Schlüsselwörtern ein

schau jedes Schlüsselwort im Text vom Satz nach

wenn es vorkommt, addiere die Beschreibung und den Programmnamen den entsprechenden Vectors

wenn der Vector mit den Beschreibungen nicht leer ist

erweitere den Kommentartext um «Es gibt eine «bung zum Thema»

für jedes Element des Vector mit den Beschreibungen

erweitere Kommentartext um: <a href={Adresse+Perlprogrammname}>{Beschreibung}</a>

wenn noch weitere Elemente vorhanden sind, erweitere Kommentartext um « und zum Thema «

sonst setze einen Punkt am Schluss

#### 4. Erzeugung der Kommentarzeilen (EdgeData) und ggf. Kommentarboxen

für jeden Satz (somit auch für jeden Kommentartext):

wenn Satz kommentiert ist, überspringe

sonst: prüfe, ob letzte Stepbox ausgehende Verbindungen zu einer Kommentarbox hat

wenn ja, erzeuge in dieser Kommentarbox eine neue Kommentarzeile

wenn nein, erzeuge eine neue Kommentarbox mit der Kommentarzeile

**Manuelle Kommentare aus «Lemma»**

- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |30:eines Satzes.  
Kommentar: |40:Während der erste Teil der Antwort noch richtig ist, stimmt der zweite überhaupt nicht. Denn ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |31:einer Phrase.  
Kommentar: |40:Während der erste Teil der Antwort noch richtig ist, stimmt der zweite überhaupt nicht. Denn ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |32:einer Wortform.  
Kommentar: |3:Korrekt. Man nennt ein Lemma auch Nennform oder Basisform. Es leitet beispielsweise einen Eintrag in einem Wörterbuch ein.  
Punkte: 15
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |33:eines Wortes.  
Kommentar: |38:Richtig. Man nennt ein Lemma auch Nennform oder Basisform eines Wortes oder genauer einer Wortform. Es leitet beispielsweise einen Eintrag in einem Wörterbuch ein.  
Punkte: 14
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |30:eines Satzes.  
Kommentar: |40:Während der erste Teil der Antwort noch richtig ist, stimmt der zweite überhaupt nicht. Denn ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |31:einer Phrase.  
Kommentar: |40:Während der erste Teil der Antwort noch richtig ist, stimmt der zweite überhaupt nicht. Denn ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |32:einer Wortform.  
Kommentar: |5:Korrekt. Man nennt ein Lemma auch Nennform oder Grundform. Es leitet beispielsweise einen Eintrag in einem Wörterbuch ein. Mit dem Begriff „Basis“ ist Vorsicht geboten, denn er ist relativ und kann für völlig unterschiedliche Zwecke gebraucht werden. So ist auch der Stamm die Basis für Flexionsprozesse oder die Wurzel die Basis für Derivationsprozesse.  
Punkte: 15
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |33:eines Wortes.  
Kommentar: |39:Korrekt. Man nennt ein Lemma auch Nennform oder Grundform eines Wortes oder genauer einer Wortform. Es leitet beispielsweise einen Eintrag in einem Wörterbuch ein. Mit dem Begriff „Basis“ ist Vorsicht geboten, denn er ist relativ und kann für völlig unterschiedliche Zwecke gebraucht werden. So ist auch der Stamm die Basis für Flexionsprozesse oder die Wurzel die Basis für Derivationsprozesse.  
Punkte: 14
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |30:eines Satzes.  
Kommentar: |40:Während der erste Teil der Antwort noch richtig ist, stimmt der zweite überhaupt nicht. Denn ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |31:einer Phrase.  
Kommentar: |40:Während der erste Teil der Antwort noch richtig ist, stimmt der zweite überhaupt nicht. Denn ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |32:einer Wortform.  
Kommentar: |7:Korrekt. Man nennt ein Lemma auch Grundform oder Basisform. Es leitet beispielsweise einen Eintrag in einem Wörterbuch ein.  
Punkte: 15
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |33:eines Wortes.  
Kommentar: |37:Richtig. Man nennt ein Lemma auch Grundform oder Basisform eines Wortes oder genauer einer Wortform. Es leitet beispielsweise einen Eintrag in einem Wörterbuch ein.  
Punkte: 14
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |30:eines Satzes.  
Kommentar: |42:Das ist total falsch. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Eine Wurzel (auch Kern genannt) hingegen ist das, was nach der Entfernung sämtlicher Affixe (Flexions- und Derivationsaffixe) von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Wurzel: ‚stell“). Eine Wurzel von Sätzen oder Phrasen gibt es gar nicht.  
Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |31:einer Phrase.  
Kommentar: |42:Das ist total falsch. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Eine Wurzel (auch Kern genannt) hingegen ist das, was nach der Entfernung sämtlicher Affixe (Flexions- und Derivationsaffixe) von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Wurzel: ‚stell“). Eine Wurzel von Sätzen oder Phrasen gibt es gar nicht.  
Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |32:einer Wortform.  
Kommentar: |9:Nein. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Eine Wurzel (auch Kern genannt) hingegen ist das, was nach der Entfernung sämtlicher Affixe (Flexions- und Derivationsaffixe) von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Wurzel: ‚stell“; Lemma: ‚herstellen“).  
Punkte: 11
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |33:eines Wortes.  
Kommentar: |43:Nein. Ein Lemma ist die Grundform (auch Nennform oder Basisform) eines Wortes oder genauer einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Eine Wurzel (auch Kern genannt) hingegen ist das, was nach der Entfernung sämtlicher Affixe (Flexions- und Derivationsaffixe) von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Wurzel: ‚stell“; Lemma: ‚herstellen“).  
Punkte: 10
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |30:eines Satzes.  
Kommentar: |41:Das ist total falsch. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Im Gegensatz dazu ist der Stamm einer Wortform das, was nach der Entfernung sämtlicher Flexionsaffixe von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Stamm: ‚herstell“; Lemma: ‚herstellen“). Einen Stamm von Sätzen oder Phrasen gibt es gar nicht.  
Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |31:einer Phrase.  
Kommentar: |41:Das ist total falsch. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Im Gegensatz dazu ist der Stamm einer Wortform das, was nach der Entfernung sämtlicher Flexionsaffixe von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Stamm: ‚herstell“; Lemma: ‚herstellen“). Einen Stamm von Sätzen oder Phrasen gibt es gar nicht.  
Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |32:einer Wortform.  
Kommentar: |41:Das ist total falsch. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Im Gegensatz dazu ist der Stamm einer Wortform das, was nach der Entfernung sämtlicher Flexionsaffixe von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Stamm: ‚herstell“; Lemma: ‚herstellen“). Einen Stamm von Sätzen oder Phrasen gibt es gar nicht.  
Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |33:eines Wortes.  
Kommentar: |41:Das ist total falsch. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Im Gegensatz dazu ist der Stamm einer Wortform das, was nach der Entfernung sämtlicher Flexionsaffixe von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen“; Stamm: ‚herstell“; Lemma: ‚herstellen“). Einen Stamm von Sätzen oder Phrasen gibt es gar nicht.  
Punkte: 4

- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |32:einer Wortform.  
Kommentar: |11:Nein. Ein Lemma ist die Grundform (auch Nennform oder Basisform) einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Der Stamm einer Wortform hingegen ist das, was nach der Entfernung sämtlicher Flexionsaffixe von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen‘; Stamm: ‚herstell‘; Lemma: ‚herstellen‘).  
Punkte: 11
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |33:eines Wortes.  
Kommentar: |44:Nein. Ein Lemma ist die Grundform (auch Nennform oder Basisform) eines Wortes oder genauer einer Wortform, also das, was in einem Wörterbuch den Eintrag einleitet. Der Stamm einer Wortform hingegen ist das, was nach der Entfernung sämtlicher Flexionsaffixe von einer Wortform übrig bleibt (z.B. Wortform: ‚herstellen‘; Stamm: ‚herstell‘; Lemma: ‚herstellen‘).  
Punkte: 10
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |34:von einem Wort übrig bleibt.  
Kommentar: |14:Nein. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Es kann allerdings der Fall eintreten, dass nach einer Trunkierung zufälligerweise genau das Lemma einer Wortform übrig bleibt.  
Punkte: 8
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |35:von einer Wortform übrig bleibt.  
Kommentar: |14:Nein. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Es kann allerdings der Fall eintreten, dass nach einer Trunkierung zufälligerweise genau das Lemma einer Wortform übrig bleibt.  
Punkte: 9
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |36:von einem Token übrig bleibt.  
Kommentar: |14:Nein. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Es kann allerdings der Fall eintreten, dass nach einer Trunkierung zufälligerweise genau das Lemma einer Wortform übrig bleibt.  
Punkte: 6
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |15:des iterativen Suffix-Stripping |34:von einem Wort übrig bleibt.  
Kommentar: |16:Das kann zwar zufälligerweise der Fall sein, gilt jedoch nicht als Definition für ‚Lemma‘. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Iteratives Suffix-Stripping meint den Prozess, bei welchem wiederholt ein Suffix von einer Wortform entfernt wird, um den Stamm zu isolieren.  
Punkte: 9
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |15:des iterativen Suffix-Stripping |35:von einer Wortform übrig bleibt.  
Kommentar: |16:Das kann zwar zufälligerweise der Fall sein, gilt jedoch nicht als Definition für ‚Lemma‘. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Iteratives Suffix-Stripping meint den Prozess, bei welchem wiederholt ein Suffix von einer Wortform entfernt wird, um den Stamm zu isolieren.  
Punkte: 10
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |15:des iterativen Suffix-Stripping |36:von einem Token übrig bleibt.  
Kommentar: |16:Das kann zwar zufälligerweise der Fall sein, gilt jedoch nicht als Definition für ‚Lemma‘. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Iteratives Suffix-Stripping meint den Prozess, bei welchem wiederholt ein Suffix von einer Wortform entfernt wird, um den Stamm zu isolieren.  
Punkte: 1
- form genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Iteratives Suffix-Stripping meint den Prozess, bei welchem wiederholt ein Suffix von einer Wortform entfernt wird, um den Stamm zu isolieren.  
Punkte: 7
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |17:der Stemmatisierung |34:von einem Wort übrig bleibt.  
Kommentar: |18:Das kann zwar zufälligerweise der Fall sein, gilt jedoch nicht als Definition für ‚Lemma‘. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Beim Stemmatisierungsprozess werden allgemein Wortformen auf den Wortstamm reduziert.  
Punkte: 10
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |17:der Stemmatisierung |35:von einer Wortform übrig bleibt.  
Kommentar: |18:Das kann zwar zufälligerweise der Fall sein, gilt jedoch nicht als Definition für ‚Lemma‘. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Beim Stemmatisierungsprozess werden allgemein Wortformen auf den Wortstamm reduziert.  
Punkte: 11
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |17:der Stemmatisierung |36:von einem Token übrig bleibt.  
Kommentar: |18:Das kann zwar zufälligerweise der Fall sein, gilt jedoch nicht als Definition für ‚Lemma‘. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet. Beim Stemmatisierungsprozess werden allgemein Wortformen auf den Wortstamm reduziert.  
Punkte: 8
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |19:des Longest Match |34:von einem Wort übrig bleibt.  
Kommentar: |20:Nein. Longest Match bezeichnet keinen Prozess, sondern eine Methode, die beispielsweise beim Prozess der Stemmatisierung angewendet werden kann. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung von einer Wortform übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 6
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |19:des Longest Match |35:von einer Wortform übrig bleibt.  
Kommentar: |20:Nein. Longest Match bezeichnet keinen Prozess, sondern eine Methode, die beispielsweise beim Prozess der Stemmatisierung angewendet werden kann. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung von einer Wortform übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 7
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |19:des Longest Match |36:von einem Token übrig bleibt.  
Kommentar: |20:Nein. Longest Match bezeichnet keinen Prozess, sondern eine Methode, die beispielsweise beim Prozess der Stemmatisierung angewendet werden kann. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was nach dem Prozess der Lemmatisierung von einer Wortform übrig bleibt und beispielsweise in einem Wörterbuch einen Eintrag einleitet.  
Punkte: 4
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |22:für Funktionswort.  
Kommentar: |23:Das ist falsch. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was in einem Wörterbuch den Eintrag einleitet. Im Gegensatz dazu sind Funktionswörter sogenannten inhaltsleer. Sie haben keine lexikalische Bedeutung und stammen aus den geschlossenen Wortklassen. Beispiele: Artikel, Pronomen, Präpositionen usw.  
Punkte: 1

- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |24:für Wortstamm.  
Kommentar: |25:Das kann zwar der Fall sein, wie bei ‚Haus‘. Dennoch dürfen Lemma und Wortstamm nicht gleich gesetzt werden. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was in einem Wörterbuch den Eintrag einleitet. Im Gegensatz dazu ist ein Stamm derjenige Teil einer Wortform, der nach Entfernung aller Flexionsaffixe übrig bleibt (z.B. Wortform: ‚herstellen‘; Stamm: ‚herstell‘; Flexionsaffix: ‚en‘; Lemma: ‚herstellen‘). Punkte: 3
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |26:für Token.  
Kommentar: |27:Falsch. Ein Lemma ist die Grundform einer Wortform (auch Nennform oder Basisform genannt), also das, was in einem Wörterbuch den Eintrag einleitet. Im linguistischen Kontext meint man mit Token dagegen jede einzelne Wortform. Allgemeiner ist ein Token irgendeine elementare Analyseeinheit. Punkte: 3
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |28:für Lexikoneintrag.  
Kommentar: |29:Das stimmt nicht ganz. Mit Lemma bezeichnet man ganz allgemein die Grundform einer Wortform (auch Nennform oder Basisform genannt). Du hast jedoch recht, dass Beispiele für Lemmas in Wörterbüchern zu finden sind. Das, was einen Eintrag in einem Wörterbuch einleitet, ist ein Lemma. Punkte: 7
- Automatische Kommentare aus «Lemma»**
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |30:eines Satzes.  
Kommentar: |45:Nicht ganz. Was ist ein Lemma? Ein Lemma ist die Grundform <b>einer Wortform.</b> Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |31:einer Phrase.  
Kommentar: |46:Nicht ganz. Was ist ein Lemma? Ein Lemma ist die Grundform <b>einer Wortform.</b> Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |32:einer Wortform.  
Kommentar: |47:Ausgezeichnet. Punkte: 15
- [0: Was ist ein Lemma? |1: Ein Lemma ist |2:die Grundform |33:eines Wortes.  
Kommentar: |48:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist die Grundform <b>einer Wortform.</b> Punkte: 14
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |30:eines Satzes.  
Kommentar: |49:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |31:einer Phrase.  
Kommentar: |50:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |32:einer Wortform.  
Kommentar: |51:Ausgezeichnet. Punkte: 15
- [0: Was ist ein Lemma? |1: Ein Lemma ist |4:die Basisform |33:eines Wortes.  
Kommentar: |52:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 14
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |30:eines Satzes.  
Kommentar: |53:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |31:einer Phrase.  
Kommentar: |54:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 8
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |32:einer Wortform.  
Kommentar: |55:Ausgezeichnet. Punkte: 15
- [0: Was ist ein Lemma? |1: Ein Lemma ist |6:die Nennform |33:eines Wortes.  
Kommentar: |56:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 14
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |30:eines Satzes.  
Kommentar: |57:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |31:einer Phrase.  
Kommentar: |58:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |32:einer Wortform.  
Kommentar: |59:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 11
- [0: Was ist ein Lemma? |1: Ein Lemma ist |8:die Wurzel |33:eines Wortes.  
Kommentar: |60:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 10
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |30:eines Satzes.  
Kommentar: |61:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |31:einer Phrase.  
Kommentar: |62:Nicht ganz. Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 4
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |32:einer Wortform.  
Kommentar: |63:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 11
- [0: Was ist ein Lemma? |1: Ein Lemma ist |10:der Stamm |33:eines Wortes.  
Kommentar: |64:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist <b>die Grundform</b> <b>einer Wortform.</b> Punkte: 10
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |34:von einem Wort übrig bleibt.  
Kommentar: |65:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stammatisierung </b> <b>von einer Wortform übrig bleibt.</b> Punkte: 8
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |35:von einer Wortform übrig bleibt.  
Kommentar: |66:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stammatisierung </b> <b>von einer Wortform übrig bleibt.</b> Punkte: 9
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |13:der Trunkierung |36:von einem Token übrig bleibt.  
Kommentar: |67:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stammatisierung </b> <b>von einer Wortform übrig bleibt.</b> Punkte: 6
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |15:des iterativen Suffix-Stripping |34:von einem Wort

- übrig bleibt.  
Kommentar: |68:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stemmatisierung </b> <b>von einer Wortform übrig bleibt.</b>  
Punkte: 9
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |15:des iterativen Suffix-Stripping |35:von einer Wortform übrig bleibt.  
Kommentar: |69:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stemmatisierung </b> von einer Wortform übrig bleibt.  
Punkte: 10
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |15:des iterativen Suffix-Stripping |36:von einem Token übrig bleibt.  
Kommentar: |70:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stemmatisierung </b> <b>von einer Wortform übrig bleibt.</b>  
Punkte: 7
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |17:der Stemmatisierung |34:von einem Wort übrig bleibt.  
Kommentar: |71:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess der Stemmatisierung <b>von einer Wortform übrig bleibt.</b>  
Punkte: 10
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |17:der Stemmatisierung |35:von einer Wortform übrig bleibt.  
Kommentar: |72:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist die Grundform einer Wortform.  
Punkte: 11
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |17:der Stemmatisierung |36:von einem Token übrig bleibt.  
Kommentar: |73:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess der Stemmatisierung <b>von einer Wortform übrig bleibt.</b>  
Punkte: 8
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |19:des Longest Match |34:von einem Wort übrig bleibt.  
Kommentar: |74:Nicht ganz. Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stemmatisierung </b> <b>von einer Wortform übrig bleibt.</b>  
Punkte: 6
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |19:des Longest Match |35:von einer Wortform übrig bleibt.  
Kommentar: |75:Nicht ganz. Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stemmatisierung </b> von einer Wortform übrig bleibt.  
Punkte: 7
- [0: Was ist ein Lemma? |12: Ein Lemma ist das, was nach dem Prozess |19:des Longest Match |36:von einem Token übrig bleibt.  
Kommentar: |76:Nicht ganz. Was ist ein Lemma? Ein Lemma ist das, was nach dem Prozess <b>der Stemmatisierung </b> <b>von einer Wortform übrig bleibt.</b>  
Punkte: 4
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |22:für Funktionswort.  
Kommentar: |77:Nicht ganz. Was ist ein Lemma? Lemma ist eine andere Bezeichnung <b>für Lexikoneintrag.</b> man hätte noch erwähnen können: Was ist ein Lemma? <b> Ein Lemma ist </b> <b>die Grundform</b> <b>einer Wortform.</b>  
Punkte: 1
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |24:für Wortstamm.  
Kommentar: |78:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Lemma ist eine andere Bezeichnung <b>für Lexikoneintrag.</b> man hätte noch erwähnen können: Was ist ein Lemma? <b> Ein Lemma ist </b> <b>die Grundform</b> <b>einer Wortform.</b>  
Punkte: 3
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |26:für Token.  
Kommentar: |79:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Lemma ist eine andere Bezeichnung <b>für Lexikoneintrag.</b> man hätte noch erwähnen können: Was ist ein Lemma? <b> Ein Lemma ist </b> <b>die Grundform</b> <b>einer Wortform.</b>  
Punkte: 3
- [0: Was ist ein Lemma? |21: Lemma ist eine andere Bezeichnung |28:für Lexikoneintrag.  
Kommentar: |80:Richtig, aber dies wäre noch besser gewesen: Was ist ein Lemma? Ein Lemma ist die Grundform einer Wortform. man hätte noch erwähnen können: Was ist ein Lemma? <b> Ein Lemma ist </b> <b>die Grundform</b> <b>einer Wortform.</b>  
Punkte: 7