

Fachübergreifende Projekte Informatik und Deutsch

- Textverarbeitung, Layouterstellung von Zeitungen
- Umgang mit und Produktion von hypermedialen Dokumenten
- Teleschreiben
- Maschinelle Sprachverarbeitung (Mustererkennung)

Anknüpfungspunkte in Themenbereichen des Deutschunterrichts

- Sprechen und Schreiben
 - Medientexte analysieren und interpretieren → vernetzte Dokumente
 - Anfertigung von Protokollen und Dokumentationen
 - Verarbeitung, Darstellung und Weitergabe von Informationen
 - Schreiben als Teamarbeit → gemeinsame Dokumente
- Umgang mit Texten und Medien
 - Kenntnisse zu, und Umgang mit Texten der Massenmedien sowie der Informations- und Kommunikationstechnologien
 - Informationen hinzufügen, reduzieren, ersetzen, klassifizieren
 - Texte auf eine andere Sprachebene bringen
- Reflexion über Sprache
 - Strukturen der Sprache als System und Funktionen ihres Gebrauchs
 - Veränderungen im Denken, Wahrnehmen und Kommunizieren durch die Informations- und Kommunikationstechnologien

Übersetzung der alten in die neue Rechtschreibung

Fachübergreifendes Projekt zur Modellbildung im Informatikunterricht

"Die deutsche Rechtschreibung hat sich über Jahrhunderte hinweg entwickelt, nicht regellos, aber auch nicht einem systematischen Konzept folgend. Einheitlichkeit in der Schreibung gibt es im deutschen Sprachraum erst seit der Rechtschreibkonferenz von 1901 bzw. seit der an sie anschließenden amtlichen Regelung der Schreibung durch die Regierungen Deutschlands, Österreichs und der Schweiz im Jahre 1902. Diese Einheitlichkeit, die heute weitgehend durch den Duden (Band1: Rechtschreibung der deutschen Sprache), in Österreich auch durch das Österreichische Wörterbuch gesichert wird, ist 1901 über Kompromisse unter konkurrierenden Regelungen und Schreibvarianten zustande gekommen – oft auf Kosten von Systematik und Einfachheit. Und manches, was an Entscheidungen in der Zeit danach (vor allem durch Einzelfallregelungen) hinzugekommen ist, hat die Erlernbarkeit der Rechtschreibung eher erschwert als erleichtert. Wie die Dinge liegen, sind selbst geübte Schreiber nicht immer in der Lage, allen Feinheiten der deutschen Rechtschreibregelung gerecht zu werden. Hier liegt der Grund dafür, dass in angemessenen Abständen versucht werden muss, Systematisierungen vorzunehmen und das Regelwerk überschaubarer und leichter handhabbar zu machen." [Duden]

"Bis auf wenige Ausnahmen, die tiefergehend untersucht werden müßten, bringt die Reform weder Transparenz noch Einfachheit in Rechtschreibung und Grammatik. Eher ist sie nutzlos, inkonsequent, undurchsichtig, verwirrend und birgt die Gefahr von Unsicherheit und neuer Fehlerquellen." [Langhans]

Wie auch immer man zur Rechtschreibreform steht, wenn sie endgültig angenommen ist, müssen tausende von Bücher in der alten Rechtschreibung (insbesondere Schulbücher) in die neue Rechtschreibung übersetzt werden.

Problemstellung:

???

Materialien:

Arbeitsanleitung

Wörterliste

Rechtschreibreform <http://www.wirtschaftsinformatik.de/wi/diverse/rechtschreibreform.htm>

Literatur zu den informatischen Grundlagen

Hinweise für den Lehrenden (inkl. Einordnung des Themas in den Unterricht)

Problemstellung:

Es soll untersucht werden, inwieweit die Übersetzung der alten in die neue Rechtschreibung automatisiert werden kann. Hierzu sind effiziente, automatisierbare Modelle für eingeschränkte Probleme zu entwerfen, zu konstruieren und zu beurteilen.

Überlegung:

Rechtschreibung beruht häufig auf der Aussprache und Bedeutung der Wörter

→ Automatisierung schwierig

Alternative:

komplette Wörter aus der alten in die Neue Rechtschreibung mit einem endlichen Automaten übersetzen

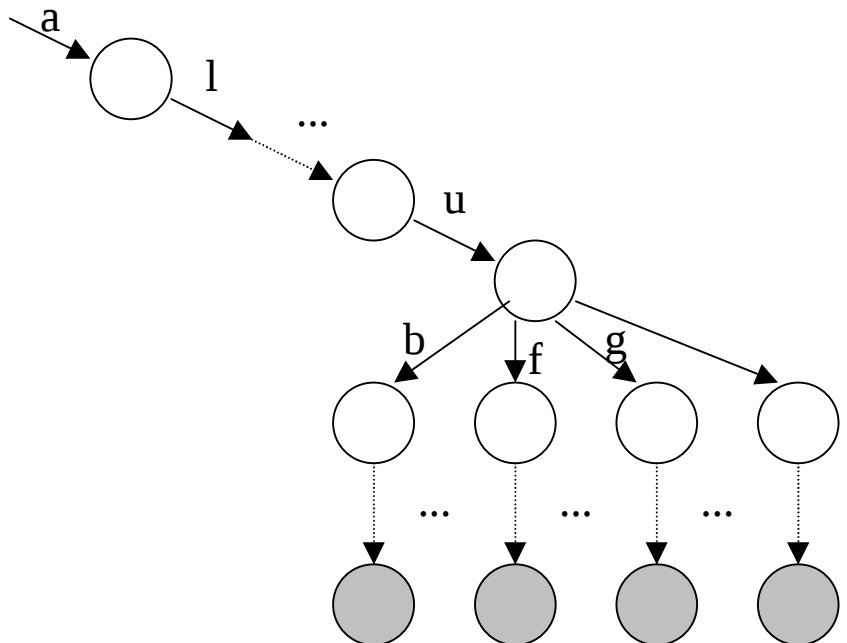
Lösungsansatz:

1. Ziel: (Scanner erzeugen):

(exemplarische) Wortliste

Quellwort	Zielwort
allzu sehr	allzu sehr
allzubald	allzu bald
allzufrüh	allzu früh
allzugern	allzu gern

- zeichenweise einlesen
- für jedes Wort wird ein Pfad in einer Datenstruktur erzeugt.
- der Knoten mit dem letzten Zeichen des Wortes bekommt ein Finalattribut
- identische Wortanfänge nutzen anfangs die gleichen Pfade, führen jedoch zu unterschiedlichen Finalknoten



Anm. die erzeugte Datenstruktur soll speicherbar sein.

2. Ziel ("lexikalische Analyse")

Der zu bearbeitende Text wird aus einer Datei zeichenweise eingelesen und in der Datenstruktur wird der entsprechende Pfad abgearbeitet.

Die Ausgaben für die akzeptierten Wörter werden in eine zweite Datei ausgegeben.

Folgende Situationen können beispielsweise im generierten (endlichen) Automaten auftreten:

es wurde ein Zeichen eingelesen für das nachfolgend keine Kante existiert
=>

WENN: der aktuelle Knoten kein Finalknoten

DANN unbekanntes Wort bis zum Ende lesen und zur Wurzel gehen

SONST

=> WENN: es handelt sich um ein Wörtertrennsymbol

DANN: gelesenes Wort ersetzen und zur Wurzel gehen

SONST: => Rechtschreibfehler

...

Problem?

Wie verhält sich der Automat, wenn der Text teilweise in der Neuen Rechtschreibung vorliegt.

Anm.

- Scanner können sehr umfangreich werden. Die Informatik kennt jedoch Verfahren, um nicht-deterministische und deterministische endliche Automaten derart ineinander umzuwandeln, dass sie minimiert werden [s. Hopcroft, 13ff].
- Am erstellten Automaten können verschiedene Optimierungsmöglichkeiten und evtl. Probleme der Mehrdeutigkeit aufgezeigt werden.
- Das Durchsuchen der gegebenen Wortliste wäre nicht effizient genug.
- Die Effizienz des endlichen Automaten kann/sollte gesteigert werden, indem Fallunterscheidungen mit einer großen Anzahl von Alternativen mittels binären Suchens behandelt werden.

Literatur:

- [Duden] http://www.duden.de/rechtschreibung/recht_frame.html
- [Hopcroft] Hopcroft, J.E.: "Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie". 2. Aufl., Addison-Wesley, Bonn 1990
- [InfoDuden] Schülerduden Informatik
- [Kastens] Kasten, U.: "Übersetzerbau". Handbuch der Informatik Bd. 3.3. Oldenbourg Verlag
- [Langhans] <http://home.t-online.de/home/j.langhans/reform.htm>
- [CU28] "Erweiterte Kulturtechnik: Schreiben". Computer+Unterricht, Heft 28, Friedrich Verlag, 1997