

6. Übungsblatt

Ausgabe: 29.11.16

Abgabe: 09.12.16, 12:00 Uhr

Haskell nach Brexit und Trump

Da die Briten und US-Amerikaner sich vom guten alten *Kontinent* abwenden werden, hat die Europäische Kommission beschlossen, bis 2050 Englisch als Verkehrssprache der EU abzuschaffen, und dies nicht nur in Protokollen, Gesetzen und Verlautbarungen, sondern auch in technischen Dokumenten und Programmen.

Also müssen alle Programme von Anglizismen befreit werden und in die verbleibenden 23 Verkehrssprachen der EU übersetzt werden. Bis 2025 sollen das zunächst nur für Schlüsselwörter geschehen, bis 2035 auch für Namen in Bibliotheken und ab 2050 dann für *alle* Namen in Programmen.

Sie sollen diesen Übergang für Programme automatisieren. Dazu entwickeln Sie ein Haskell-Programm (im bis 2025 noch erlaubten angelsächsischen Dialekt), dass alle Bezeichner (*identifier*) in Programmen mit Hilfe eines Wörterbuchs übersetzt.

6.1 Ein abstrakter Datentyp für Wörterbücher

10 Punkte

Für endliche Abbildungen $\text{Map } [\alpha] \beta$ kann eine Realisierung als balancierter Baum ineffizient sein, weil beim Einfügen und Suchen logarithmisch oft Listen verglichen werden müssen. In diesem Fall ist es günstiger, sich die sequenzielle Struktur der Schlüssel zu Nutze zu machen und einen *Präfixbaum* oder *Trie* (auch bekannt als *digital search tree*) zu benutzen.

Ein *Trie* ist ein n -stelliger Baum, dessen Knoten mit Werten der Bildmenge β markiert sein können, und dessen Kanten mit Elementwerten aus α markiert sind. Der *Trie* in Abbildung 1 stellt eine $\text{Map } [\alpha] \beta$ mit $\alpha = \text{Char}$ und $\beta = \text{String}$ dar. Die Kantenmarkierungen aller Wege von der Wurzel zu markierten Knoten legen den Definitionsbereich des *Tries* fest; im Beispiel ist dies $\{\text{cap}, \text{car}, \text{cat}, \text{do}, \text{dog}\}$.

Um den einer Zeichenkette w zugeordneten Wert zu finden, muss man von der Wurzel aus einen mit w markierten Weg suchen. Gibt es so einen Weg nicht, oder ist sein Zielknoten nicht markiert, ist w nicht im Definitionsbereich des *Tries* enthalten, wie z.B. "cart" oder "ca" in Abbildung 1.

Wenn wir *Tries* als algebraischen Datentypen $\text{Trie } \alpha \beta$ darstellen, enthalten die Knoten Einträge vom Typ *Maybe* β . In Abbildung 1 enthalten die leeren Knoten *Nothing* und die anderen *Just* "Kappe" usw. Die Unterbäume sollen als endliche Abbildungen von α auf *Tries* dargestellt werden.¹ Die Wurzel in Abbildung 1 hat bspw. den Definitionsbereich $[\text{'c'}, \text{'d'}]$. Für die Blätter der *Tries* sind diese Abbildungen gänzlich undefiniert.

Implementieren Sie einen abstrakten Datentypen *Trie* mit den in Abbildung 2 angegebenen Operationen, deren Funktionalitäten Ihnen aus Übungsblatt 5 bekannt sein sollten, sowie einer geeigneten Instanz für die Typklasse *Eq*.

6.2 Haskellprogramme übersetzen

7 Punkte

Implementieren Sie nun ein Haskell-Modul *TranslateHaskell* mit folgenden Funktionen:

`translate :: Trie Char String → String → String`

`germanHaskell :: String → String`

Das erste Argument von `translate` enthält ein Wörterbuch von Bezeichnern. Jedes Auftreten eines Bezeichners x im zweiten Argument, (dem *Originalprogramm*) soll im Ergebnis (dem *Zielprogramm*) durch den entsprechenden Bezeichner d der Zielsprache ersetzt werden. Dabei sind einige Besonderheiten zu beachten:

¹Hierzu können Sie den Typ `Data.Map` aus der Haskell-Standardbibliothek benutzen.

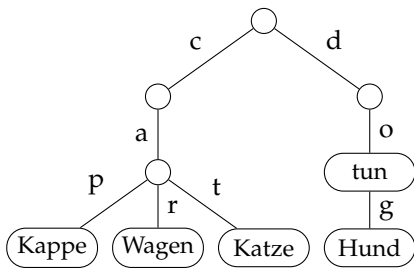


Abbildung 1: Ein Präfixbaum (Trie)

```

empty  :: Trie α β
null   :: Trie α β → Bool
insert :: Ord α ⇒ [α] → β → Trie α β → Trie α β
delete :: Ord α ⇒ [α] → Trie α β → Trie α β
lookup :: Ord α ⇒ [α] → Trie α β → Maybe β
(!)    :: Ord α ⇒ Trie α β → [α] → β
toList :: Trie α β → [[α], β]
keys   :: Trie α β → [[α]]
fromList :: Ord α ⇒ [[α], β] → Trie α β

```

Abbildung 2: Schnittstelle von Trie

1. Als Bezeichner sollen Zeichenketten aus Buchstaben erkannt werden.²
2. Es werden *maximale* Bezeichner erkannt, d.h., davor und dahinter müssen andere Zeichen als Buchstaben stehen.
3. Bezeichner, die im Wörterbuch fehlen, bleiben unverändert.

Die Signatur von `translate` erlaubt es, beliebige Wörterbücher für beliebige Programmiersprachen anzugeben und somit zwischen beliebigen nationalen Dialekten dieser Programmiersprache hin und her zu übersetzen, beispielsweise von Litauisch nach Ungarisch. Sie sollen mit `translate` aber nur die Funktion `germanHaskell` implementieren und testen, die Schlüsselwörter englischen Haskell's ins Deutsche übersetzt.³ Stringlitterale und Kommentare brauchen nicht gesondert behandelt (d.h. vor der Übersetzung geschützt) zu werden.

6.3 Robustheit

3 Punkte

Sei `p` ein (angelsächsisches) Haskell Programm und `p' = germanHaskell p` seine Übersetzung in kontinentales Haskell. Dann kann `p'` z.B. vom deutschen Haskell-System nicht verarbeitet werden, wenn im Originalprogramm `p` deutsche Schlüsselwörter als Bezeichner benutzt wurden, etwa "if wenn then dann else sonst".

Dieses Problem sollen Sie beheben, indem Sie in `p` alle kontinentalen Schlüsselwörter in die angelsächsischen zurück übersetzen. Das obige Beispiel würde dann so übersetzt: "wenn if dann then sonst else".

Hinweise:

- Beim Zerstückeln von Programmen in Listen von Bezeichnern und Sonstigem helfen die Funktionen

```

span    :: (α → Bool) → [α] → ([α], [α])
break   :: (α → Bool) → [α] → ([α], [α])
isLetter :: Char → Bool                — aus Data.Char

```

- Die Datei `Main.hs` enthält eine Funktion `run`, die mit einem Dateinamen als Argument aufgerufen wird, diese Datei liest, auf ihrem Inhalt `germanHaskell` anwendet, und das Ergebnis ausgibt:

```

*Main> Main.run "Main.hs"
modul Main wobei

importiere System.Environment
importiere GermanHaskell
....

```

? Verständnisfragen

1. Was ist ein abstrakter Datentyp (ADT)?
2. Was sind Unterschiede und Gemeinsamkeiten zwischen ADTs und Objekten, wie wir sie aus Sprachen wie Java kennen?
3. Wozu dienen Module in Haskell?

²Erst im Wintersemester 2034/2035 müssen sich Ihre Nachfolger dann mit allgemeinen Bezeichnern wie `Willi_2` auseinandersetzen.

³Eine entsprechende Liste `en_de` von Schlüsselwortpaaren ist vorgegeben.