Programmierkurs Python II, Sommersemester 2010, Übungsblatt 1

Aufgabe 1: Bäume implementieren

Implementiere eine Klasse für Bäume (siehe Folien). Die Klasse soll mindestens die folgenden Methoden implementieren:

- `__repr__(self)` konvertiert einen Baum in einen String.
- `__iter__(self)` liefert einen Iterator über die Knoten im Baum.
- `leaves(self)` liefert einen Iterator über die Blätter des Baums.
- `depth(self)` berechnet die Tiefe des Baums.

Aufgabe 2: Baumausdrücke parsen

Implementiere einen Parser für Baumausdrücke. Die Funktion `tree` und `trees` können von den Folien übernommen werden.

Der Tokenisierer soll für eine Zeichenkette als Eingabe einen Iterator über die Tokens liefern:

```
>>> list(tokenize("(S (NP Peter) (VP schläft))"))
['(', 'S', '(', 'NP', 'Peter', ')', '(', 'VP', 'schläft', ')', ')']
```

Der Parser soll als Iterator implementiert werden.

```
class Parser:
    def __init__(self, tokens):
        self.tokens = tokens
    def __iter__(self):
        return self
    def __next__(self):
        return tree(next(self.tokens), self.tokens) # siehe Folien
```

```
>>> t = tokenize("(S (NP Peter) (VP schläft)) (S (NP Er) (VP schnarcht))")
>>> p = Parser(t)
>>> next(p)
Tree('S', [Tree('NP', [Tree('Peter', [])]), Tree('VP', [Tree('schläft', [])])])
>>> next(p)
Tree('S', [Tree('NP', [Tree('Er', [])]), Tree('VP', [Tree('schnarcht', [])])])
>>> next(p)
StopIteration
```

Erweitere die Implementierung der Funktionen `tree` bzw. `trees` so, dass Fehler in der Eingabe korrekt behandelt werden (beispielsweise durch Werfen einer Ausnahme).

Aufgabe 3: Nichtterminalsymbole zählen

Auf der Homepage findet sich eine kleine Beispieldatei mit einer Liste von (zwei) Bäumen. Implementiere ein Programm, dass die Bäume aus der Datei einliest und zählt, wie häufig welche Nichtterminalsymbole vorkommen. Als Nichtterminalsymbol sollen alle Etiketten von Nicht-Blättern zählen.

Versuche, die Implementierung so zu gestalten, dass die Eingabedatei möglichst nur Zeichen- oder Zeilenweise eingelesen wird.

Abgabe bis Dienstag, 2010-05-04, 08:30 Uhr