

Ausgabe: 15.12.2017

Abgabe: 17.01.2018

## Hausaufgabe: Spamfilter für das Internet der Dinge

Mit Ihrem erfolgreichen Start-up haben Sie ein Internet-of-Things-Gadget entwickelt, das E-Mails empfangen kann. Unpraktischerweise stellt sich heraus, dass Ihr Gadget jede Menge Spam bekommt. Deshalb benötigen Sie nun einen Spamfilter. Da Ihr Gadget einen MIPS-Prozessor verwendet, soll der Spamfilter in MIPS-Assembler geschrieben werden. Der Spamfilter durchsucht den Text einer E-Mail nach Wörtern aus einer Liste. Jedes dieser Wörter hat ein festgelegtes Gewicht. Ist die Summe der Gewichte über einer Schwelle, so soll die E-Mail als Spam markiert werden.

### Hinweise zur Bearbeitung:

- Ihre Abgabe muss unter der QtSPIM-Installation auf den TU-Berlin-Rechnerpools (z.B. TEL 106 und TEL 206) lauffähig sein. Ihr Lösungsweg muss nachvollziehbar sein; z.B. durch Kommentare.
- Für alle Aufgaben erhalten Sie eine Vorgabe, die Teil der Aufgabenstellung ist. Die Teilaufgaben sind einzeln lösbar.
- Sie erhalten zu jeder Aufgabe ein Python-Script mit dem Sie korrekte Ausgaben generieren und Ihre eigenen Ergebnisse überprüfen können.
- Nutzen Sie die (Qt)SPIM-Dokumentation<sup>1</sup>. Beachten Sie vorallem das Dokument *Appendix A: Assemblers, Linkers, and the SPIM Simulator*.

### 1. E-Mail einlesen (4 Punkte)

Das Format einer E-Mail ist in RFC 5322 definiert. Eine Zusammenfassung, die die relevanten Teile abdeckt, finden Sie in der Wikipedia<sup>2</sup>. Eine E-Mail ist in ASCII kodiert und besteht aus Header und Body. Im ersten Schritt muss die E-Mail in *Header* und *Body* zerlegen, damit danach der *Body* durchsucht werden kann.

Implementieren Sie Routine `parse_email`, die als Rückgabewerte das Offset von *Subject* und *Body* zurück gibt.

### 2. Spamfilter (12 Punkte)

Um zu erkennen, ob eine E-Mail als Spam zu klassifizieren ist, vergleicht der Spamfilter alle Wörter im Text der E-Mail mit einer Liste von *Bad Words*. Diese Wörter haben jeweils ein Gewicht, das die Wahrscheinlichkeit, dass es sich um ein Wort aus einer Spam-E-Mail handelt, abbildet.

Implementieren Sie die Routine `spamfilter`, die einen Spamscore für eine E-Mail berechnet.

### 3. E-Mail ausgeben (4 Punkte)

Der Spamfilter gibt alle E-Mails wieder aus. Falls eine E-Mail als Spam identifiziert wurde, soll dem Betreff der E-Mail die Markierung "[SPAM]" vorangestellt werden.

Implementieren Sie die Routine `print_email`, die die E-Mail wieder ausgibt und dabei ggf. verändert.

<sup>1</sup><http://spimsimulator.sourceforge.net/further.html>

<sup>2</sup>[https://de.wikipedia.org/wiki/Header\\_\(E-Mail\)](https://de.wikipedia.org/wiki/Header_(E-Mail))