

Dr. Jürg M. Stettbacher

Margrit Rainer Strasse 12a
CH-8050 Zürich

Telefon: +41 43 299 57 23
Fax: +41 43 299 57 25
E-Mail: dsp@stettbacher.ch

LZ77 Encoder und Decoder

Praktikum

Version 2.00
2013-06-12

Zusammenfassung: In diesem Praktikum geht es darum, je einen LZ77 Encoder und einen LZ77 Decoder zu schreiben. Es soll also möglich sein, gegebene Daten aus einer Datei in LZ77 Tags umzuwandeln und anschliessend daraus die ursprünglichen Daten wieder herzustellen.

Inhaltsverzeichnis

1	Einleitung	2
2	Dateien	2
3	Encoder Aufgabe	3
4	Decoder Aufgabe	3

1 Einleitung

Der LZ77 Encoder ist ein typischer Quellenencoder. Er wandelt unterschiedlich lange Sequenzen von Quellensymbolen, zum Beispiel ASCII-Zeichen, in Tags gleicher Länge um. Dabei werden insbesondere ähnliche Symbolabfolgen der Quelle effizient zusammen gefasst. Der LZ77 Encoder verwendet zu diesem Zweck ein sog. *sliding Window*, also einen Such- und einen Vorschabuffer. Es wird eine möglichst lange Übereinstimmung zwischen den aktuellen Zeichen im Vorschabuffer und einer beliebigen Stelle im Suchbuffer gesucht. Anschliessend wird die betreffende Stelle im Suchbuffer durch einen Token beschrieben.

In diesem Praktikum haben wir verschiedene Datenquellen in der Form von Dateien zur Verfügung. Es handelt sich um ASCII-Dateien (*.txt) und jedes ASCII-Zeichen daraus stellt ein Symbol dar. Gesucht ist ein LZ77 Encoder und Decoder, so dass diese Daten codiert und anschliessend wieder decodiert werden können.

2 Dateien

Die folgenden Dateien stehen Ihnen für das Bearbeiten des Praktikums zur Verfügung:

- *lz77-enc_template.php* (Encoder)

- *lz77-dec_template.php* (Decoder)
- *data_1.txt* bis *data_6.txt* und *deutsch.txt* (Testdaten)

3 Encoder Aufgabe

Schreiben Sie zuerst ein PHP-Skript *lz77-enc.php*, das eine bestimmte Datei öffnet und deren Inhalt in LZ77 Tags umwandelt. Der Name der ASCII-Datei soll auf der Kommandozeile übergeben werden, so dass der Aufruf folgendermassen aussieht.

```
> php lz77-enc.php data.txt
```

Das Resultat soll in eine neue Datei mit dem Namen *data.txt.lz77* geschrieben werden. Beachten Sie, dass wir uns hier nicht die Mühe machen wollen, die Tags binär zu codieren. Statt dessen schreiben wir sie einfach im für uns lesbaren ASCII-Format in die Datei. Damit Encoder und Decoder sicher mit den selben Buffergrössen arbeiten, schreiben wir die beiden Grössenangaben (in Symbolen pro Buffer) ganz am Anfang in die LZ77 Datei.

Sie erhalten eine Vorlage *lz77-enc_template.php*, in welcher einige Funktionsblöcke bereits vorhanden sind. Sie müssen nur an den bezeichneten Stellen ergänzen. Für die Erzeugung der LZ77 Tags gehen Sie so vor:

- Suche Sie die längste Übereinstimmung von Such- und Vorschabuffer.
- Erzeugen Sie den Token.
- Schieben Sie die verarbeiteten Quellensymbole vom Vorschau- in den Suchbuffer und füllen Sie den Vorschabuffer mit neuen Quellensymbolen auf. (In anderen Worten: Schieben Sie das sliding Window weiter.)

Diese Aufgabe ist zu erfüllen, und wenn nötig im Selbststudium fertig zu stellen.

4 Decoder Aufgabe

Schreiben Sie ein zweites PHP Skript, nämlich *lz77-dec.php*, welches die zuvor erzeugte Datei *data.txt.lz77* wieder decodiert. Die decodierte Datei soll *data.txt.lz77.dec* heissen, so dass sie mit der ursprünglichen Datei *data.txt* verglichen werden kann.

Als Vorlage erhalten Sie die Datei *lz77-dec_template.php*, in welcher einige Funktionsblöcke bereits vorhanden sind. Sie müssen nur an den bezeichneten Stellen ergänzen.

Diese Aufgabe ist zu erfüllen, und wenn nötig im Selbststudium fertig zu stellen.