patgen selbst kompilieren

Die Wortliste der deutschen Trennmustermannschaft ist sehr groß. Die originale Version von `patgen` ohne Patches reicht keinesfalls aus, um deutsche Trennmuster zu erstellen. TeXLive kommt bereits seit 2006 mit um den Faktor 10 vergrößerten Werten für `trie_size` und `triec_size`, doch reicht das immer noch nicht, wenn man mit der aktuellen Liste experimentieren will – die beiden Arrays müssen nochmals um den Faktor 10 vergrößert werden, was in TeXLive 2019 geschehen ist.

Will man mit Primär- und Sekundärtrennmuster experimentieren, sind zusätzliche Anpassungen notwendig, und man muss daher `patgen` selbst kompilieren. Hier eine Anleitung, wie man das unter einem Unix-ähnlichen Betriebssystem wie GNU/Linux oder MacOS macht.

Quellcode holen

[Der hier beschriebene Weg braucht rund 500MByte an Festplattenspeicher, um eine Datei mit einer Größe von weniger als 100kByte zu erzeugen – das ist natürlich mit Kanonen auf Spatzen schießen, jedoch wird keine zusätzliche Handarbeit benötigt, was das ganze Unterfangen enorm erleichtert.]

Die schnellste Methode ist, mittels `rsync` nur den Inhalt des Verzeichnisses `build/source` des TeXLive-Repositoriums zu holen.

```
mkdir patgen-build
cd patgen-build
mkdir source
cd source
rsync -vaz --delete --exclude=.svn tug.org::tldevsrc/Build/source/ .
cd ..
```

Insgesamt werden rund 90MByte komprimiert übertragen, was rund 430MByte Daten auf der Festplatte ergibt.

Anpassungen für Primär- und Sekundärtrennstellenmuster vornehmen

Falls Trennmuster erzeugt werden sollen, die nicht alle Trennstellen, sondern nur solche an Wortfugen der geringsten Bindungsstärke (Primärtrennstellen) oder der geringsten und zweitgeringsten Bindungsstärke (Sekundärtrennstellen) berücksichtigen, muss der Quellcode von `patgen` bearbeitet werden. Der Grund dafür ist, dass bei der Mustererzeugung längere Wortteile betrachtet werden müssen, als es mit den voreingestellten Parametern möglich ist.

In der Datei `source/texk/web2c/patgen.ch` müssen die Zeilen

```
@!trie_size=10000000; {space for pattern trie}
@!triec_size=5000000; {space for pattern count trie, must be less than
```

gesucht und beide Werte verdoppelt werden:

```
@!trie_size=20000000; {space for pattern trie}
@!triec_size=10000000; {space for pattern count trie, must be less than
```

In der Datei `source/texk/web2c/patgen.web` muss die Zeile

```
@!max_dot=15; {maximum pattern length, also maximum length of external
```

gesucht und der Wert auf 35 hinaufgesetzt werden:

```
@!max_dot=35; {maximum pattern length, also maximum length of external
```

Konfiguration und Kompilation

Wir bauen `patgen` außerhalb des Quellcode-Baumes.

```
mkdir work
cd work
```

Am einfachsten ist es, den folgenden langen Befehl mittels Cut-and-Paste in einer Datei `build.sh` zu speichern und diese dann als `sh build.sh` auszuführen.

```
../source/configure \
  --disable-all-pkgs \
  --enable-web2c \
  --disable-tex \
  --disable-etex-synctex \
  --disable-ptex \
  --disable-ptex-synctex \
  --disable-eptex \
  --disable-eptex-synctex \
  --disable-uptex \
  --disable-uptex-synctex \
  --disable-euptex \
  --disable-euptex-synctex \
  --disable-aleph \
  --disable-hitex \
  --disable-pdftex \
  --disable-pdftex-synctex \
  --disable-luatex \
  --disable-luajittex \
  --disable-luahbtex \
  --disable-luajithbtex \
  --disable-xetex \
  --disable-synctex \
  --disable-mp \
  --disable-pmp \
```

```
  --disable-upmp \
  --disable-mfluajit \
  --disable-mf \
  --disable-mf-nowin \
  --disable-mflua \
  CFLAGS="-O3" &> configure.log \
&& make &> make.log
```

Die Ausführung des obigen Befehls dauert ein paar Minuten und benötigt rund 50MByte an Festplattenspeicher; die stdout-Ausgabe wird in die Dateien `configure.log` und `make.log` umgeleitet.

Installation

Die kompilierte Datei `patgen` befindet sich im Verzeichnis `texk/web2c`. Da das Programm statisch kompiliert ist und daher von keinen DLLs abhängig ist, kann man die Datei in ein beliebiges Verzeichnis kopieren.

From:
https://wiki.dante.de/ - **DanteWiki**

Permanent link:
https://wiki.dante.de/doku.php?id=trennmuster:patgen_selbst_kompilieren

Last update: **2023/08/19 00:11**