

Performance-Analyse von Lingo

Testdurchführung

Um die Performance von Lingo zu verbessern, wird im ersten Schritt eine grobe Analyse der Zeitanteile je Modul durchgeführt. Dies soll Aufschlüsse darüber geben, in welchem Modul eine Performanceoptimierung am dringenden angesagt ist. Als Textdatei wird der Einfachheit halber das Standardwörterbuch als Eingabe benutzt. Die ersten drei Tests geben den groben Zeitrahmen wieder, den Lingo in der Standardkonfiguration benötigt.

Ab Test 4 werden die Module sukzessive auf ihren Zeitanteil bei der Abarbeitung hin untersucht.

Test	Beschreibung	Laufzeit
1	Aufruf von ruby lingo.rb de/lingo-dic.txt mit Standardkonfiguration und Bildschirmausgabe	627 sec
2	gleiche Funktion ohne Bildschirmausgaben	517 sec
3	gleiche Funktion ohne Dateiausgaben	413 sec
4	Nur Textreader Anzahl Bytes = 1734844 Anzahl Dateien = 1 Anzahl Zeilen = 62041	0,43 sec
5	+ Tokenizer Anzahl Muster HELP = 12713 Anzahl Muster NUMS = 4 Anzahl Muster OTHR = 127219 Anzahl Muster PUNC = 6 Anzahl Muster URLS = 1 Anzahl Muster WORD = 203018 Anzahl Token = 342961 Anzahl Zeilen = 62041 Objekte empfangen = 62041	9,4 sec
6	+ Abbreviator Anzahl gesuchter Abkürzungen = 62037 Objekte empfangen = 342961 TOPs empfangen = 62044 cache hits = 62028 sys-abk: source reads = 11 sys-abk: total requests = 11	15,45 sec

Test	Beschreibung	Laufzeit
7	+ Wordsearcher Anzahl gefundener Wörter = 112716 Anzahl gesuchter Wörter = 203018 Objekte empfangen = 342961 TOPs empfangen = 62044 cache hits = 146888 sys-dic: cache hits = 1858 sys-dic: data found = 51459 sys-dic: source reads = 58088 sys-dic: total requests = 59946	40,97 sec
8	+ Decomposer Komposita erkannt = 752 Komposita geprüft = 2813 Objekte empfangen = 342961 String zu kurz = 87412 TOPs empfangen = 62044 cache hits = 721 sys-dic: cache hits = 21230 sys-dic: data found = 1152 sys-dic: source reads = 21408 sys-dic: total requests = 42638	52,94 sec
9	+ Multiworder Objekte empfangen = 342961 TOPs empfangen = 62044 sys-mul: cache hits = 215289 sys-mul: source reads = 470626 sys-mul: total requests = 685915	376 sec
10	+ Sequencer Anzahl erkannter Sequenzen = 1 Objekte empfangen = 342961 TOPs empfangen = 62044	400 sec
11	+ Synonymer Anzahl erweiterter Wörter = 1310 Anzahl gefundener Synonyme = 1899 Anzahl gesuchter Wörter = 113674 Objekte empfangen = 342962 TOPs empfangen = 62044 sys-syn: cache hits = 89953 sys-syn: data found = 472 sys-syn: source reads = 49481 sys-syn: total requests = 139434	427 sec

Testanalyse

Der Zeitanteil der einzelnen Attendees zeigt sich damit in der nachfolgenden Grafik. Dies zeigt klar, dass die Performance des Multiworder als erstes optimiert werden sollte. Vorab sollte man bei Projekten abwägen, ob die Funktionalität des Multiworders zwingend notwendig, oder ob auf ihn zugunsten einer deutlich gesteigerten Performance verzichtet werden kann.

