

テキスト処理 第12回 (2007-07-24)

dq2sqレポート説明

田中哲

産業技術総合研究所

情報技術研究部門

`akr@isc.senshu-u.ac.jp`

`http://staff.aist.go.jp/tanaka-akira/textprocess-2007/`

レポート

- 文字列中にダブルクォートで括られた部分があるとして、その括りをシングルクォートに変えるメソッド `dq2sq` を実装せよ
- `dq2sq(str)`
- 実装したらユニットテストで確認してほしい
- ✖切 2007-07-24 16:20
- HIPLUS
- 拡張子が `txt` なテキストファイルがよい

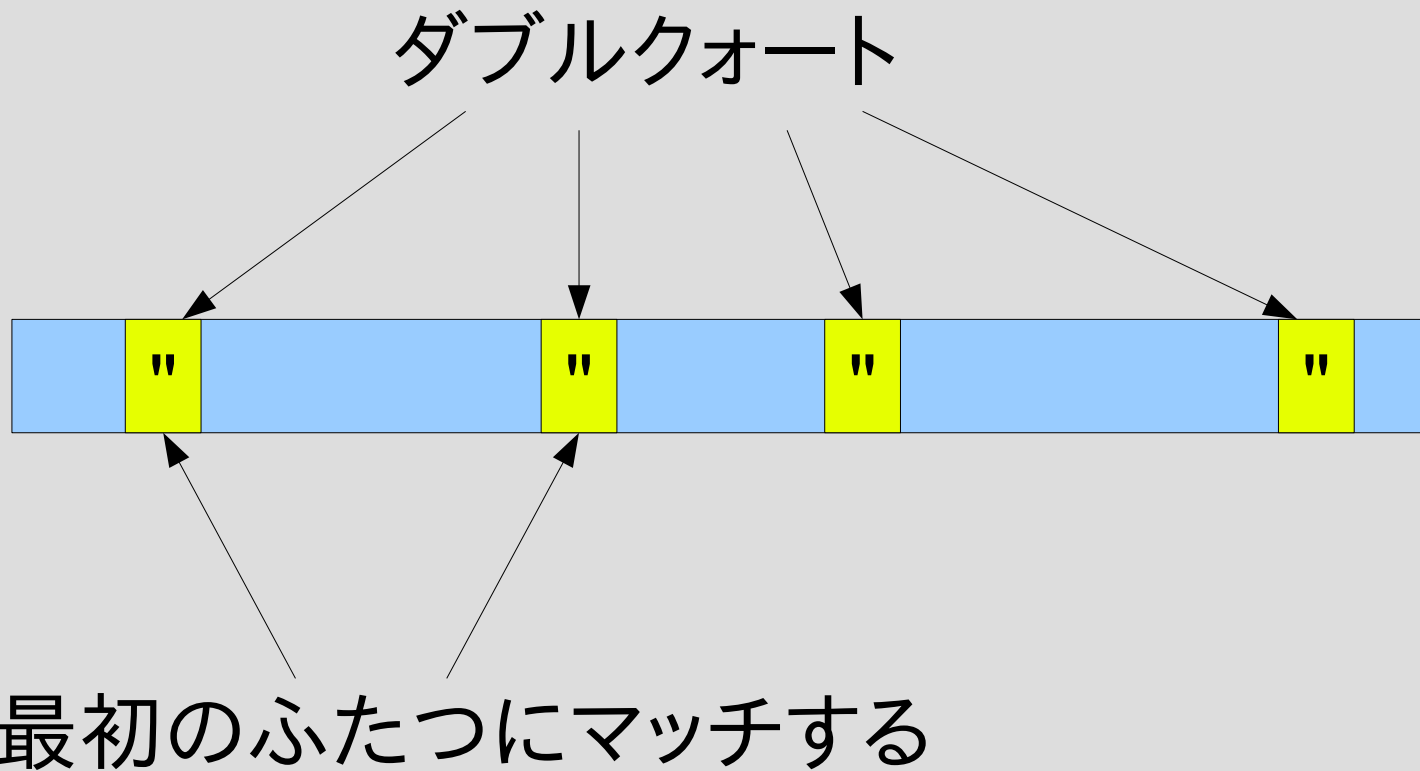
動作例

- `dq2sq('abc"def"ghi')` $\#=>$ `"abc'def'ghi"`
- `dq2sq('a"b"c"d"e')` $\#=>$ `"a'b'c¥"d¥"e"`
- `dq2sq('abc')` $\#=>$ `"abc"`

ヒント

- subst を一回使う
- ダブルクォートで括弧してあるものがふたつ以上あっても最初のひとつだけ変える
- ダブルクォートで括弧してあるものがひとつもなければなにも変えずに返す

ダブルクォートの対にマッチ



dq2sqの実装

```
def dq2sq(str)          シングルクォート、ダブルクォート、
  subst(str,           シングルクォート、3文字の並び
    [:cat, [:lit, ""],
      [:cat, [:capture, :content,
        [:rep_lazy, [:anysym]]],
        [:lit, ""]]]) { |s, h|
      "#{h[:content]}"
    }
  end
```

ダブルクォート、シングルクォートの並び

String#sub での表現

```
def dq2sq(str)
  str.sub(/"((?:.|\\n)*?)"/) { "#{$1}" }
end
```

dq2sqの中身 (1)

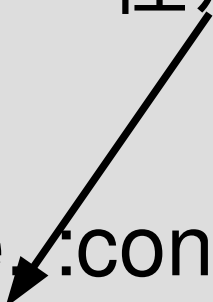
```
def dq2sq(str)
  subst(str,
    [:cat, [:lit, ""],
    [:cat, [:capture, :content,
    [:rep_lazy, [:anysym]]],
    [:lit, ""]]]) {|s, h|
    ""#{h[:content]}""
  }
end
```

ダブルクォートふたつ

dq2sqの中身 (2)

```
def dq2sq(str)
  subst(str,
    [:cat, [:lit, ""],
      [:cat, [:capture, :content,
        [:rep_lazy, [:anysym]]],
        [:lit, ""]]]) {|s, h|
    "#{h[:content]}"
  }
end
```

任意の文字列
ただしなるべく短いもの



dq2sqの中身 (3)

```
def dq2sq(str)
  subst(str,
    [:cat, [:lit, ""],
      [:cat, [:capture, :content,
        [:rep_lazy, [:anysym]]],
        [:lit, ""]]]) {|s, h|
    ""#{h[:content]}""
  }
end
```

キャプチャした文字列を使う

ざっと眺めた結果

- ダブルクォートを含む文字列の書き方
 - Cと同様に "¥" と書ける
 - シングルクォートを使えば '¥' と書ける
 - 文字列の記法は第2回の資料ですこし解説してある
- ダブルクォート間の文字列を取り出すのに `s[1...(s.size-1)]` としているものがあった
 - キャプチャを使うことを意図していたが、問題はない
 - ちなみに `s[1...-1]` ともかける
- 2回 `subst` しているものがあった
 - 想定とは異なる
 - ちゃんと動くものもあった
 - 動かないものもあった (対でないときも変えてしまう)