Hard Solution

Load Packages

```r
library(xml2)
library(stringr)
library(rmarkdown)
```

```r
z <- '
<CATALOG>
  <CD>
    <TITLE>Empire Burlesque</TITLE>
    <ARTIST>Bob Dylan</ARTIST>
    <COUNTRY>USA</COUNTRY>
    <COMPANY>Columbia</COMPANY>
    <PRICE>10.90</PRICE>
    <YEAR>1985</YEAR>
  </CD>
  <CD>
    <TITLE>Hide your heart</TITLE>
    <ARTIST>Bonnie Tylor</ARTIST>
    <COUNTRY>UK</COUNTRY>
    <COMPANY>CBS Records</COMPANY>
    <PRICE>9.90</PRICE>
    <YEAR>1988</YEAR>
  </CD>
</CATALOG>
'
```

```r
res <- rlist::list.parse(z, type='xml')
res
```

```
$CD
$CD$TITLE
[1] "Empire Burlesque"

$CD$ARTIST
[1] "Bob Dylan"
```

[1] "Bob Dylan"

CDCOUNTRY
[1] "USA"

CDCOMPANY
[1] "Columbia"

CDPRICE
[1] "10.90"

CDYEAR
[1] "1985"

$CD
CDTITLE
[1] "Hide your heart"

CDARTIST
[1] "Bonnie Tylor"

CDCOUNTRY
[1] "UK"

CDCOMPANY
[1] "CBS Records"

CDPRICE
[1] "9.90"

CDYEAR
[1] "1988"

```
doc2 <- read_xml(z)
doc2
```

{xml_document}
<CATALOG>
[1] <CD>\n <TITLE>Empire Burlesque</TITLE>\n <ARTIST>Bob

Dylan</ARTIST>\n   ...
[2] <CD>\n <TITLE>Hide your heart</TITLE>\n <ARTIST>Bonnie
Tylor</ARTIST>\n ...

extract CDs, branches of trees

```
        cds <- xml_children(doc2)
        cds
```

{xml_nodeset (2)}
[1] <CD>\n <TITLE>Empire Burlesque</TITLE>\n <ARTIST>Bob
Dylan</ARTIST>\n   ...
[2] <CD>\n <TITLE>Hide your heart</TITLE>\n <ARTIST>Bonnie
Tylor</ARTIST>\n ...

res is a list of 2, CD is a list of 6

```
        cd_list <- lapply(cds,function(cd){
          cd_data <- xml_children(cd)
          })
        cd_list
```

[[1]]
{xml_nodeset (6)}
[1] <TITLE>Empire Burlesque</TITLE>
[2] <ARTIST>Bob Dylan</ARTIST>
[3] <COUNTRY>USA</COUNTRY>
[4] <COMPANY>Columbia</COMPANY>
[5] <PRICE>10.90</PRICE>
[6] <YEAR>1985</YEAR>

[[2]]
{xml_nodeset (6)}
[1] <TITLE>Hide your heart</TITLE>
[2] <ARTIST>Bonnie Tylor</ARTIST>
[3] <COUNTRY>UK</COUNTRY>
[4] <COMPANY>CBS Records</COMPANY>
[5] <PRICE>9.90</PRICE>
[6] <YEAR>1988</YEAR>

```
[6] <YEAR>1988</YEAR>
```

```
my_parse <- function(z) {
  doc2 <- read_xml(z)
  cds <- xml_children(doc2)
  cd_list <- lapply(cds, function(cd) {
    cd_data <- xml_children(cd)
  })

  list_main <- list()
  child_one <- xml_name(cds)

  for (element1 in 1:length(child_one)) {
    cdnames <- xml_name(cd_list[[element1]])
    cdtext <- xml_text(cd_list[[element1]])
    print(cdnames)
    print(cdtext)

    list_cd <- list()
    for (element2 in 1:length(cdnames)) {
      list_cd[cdnames[element2]] <- cdtext[element2]
    }

    list_main[[element1]] <- list_cd
  }
  names(list_main) <- child_one
  return(list_main)
}
```

```
res2 <- my_parse(z)
```

```
[1] "TITLE"    "ARTIST"   "COUNTRY" "COMPANY" "PRICE"    "YEAR"
[1] "Empire Burlesque" "Bob Dylan"          "USA"
"Columbia"
[5] "10.90"            "1985"
[1] "TITLE"    "ARTIST"   "COUNTRY" "COMPANY" "PRICE"    "YEAR"
[1] "Hide your heart" "Bonnie Tylor"    "UK"               "CBS
Records"
[5] "9.90"            "1988"
```

```r
identical(res, res2)
```

```
[1] TRUE
```