Queries in MongoDB

I'm trying to use `rmongodb` to fetch information from a MongoDB database for further processing in R. However, I have some difficulties to really get started. This one works:

```
cursor <- mongo.find(mongo, "people", query=list(last.name="Smith", first.name="John"
                 fields=list(address=1L, age=1L))
while (mongo.cursor.next(cursor)){
  print(mongo.cursor.value(cursor))}
```

Now, what if I want to find people whose first name is either "John" or "Bob" or "Catherine"? I tried `query=list(last.name="Smith", first.name=c(John, Bob, Catherine))` but this didn't work out. Replacing `=` with `%` didn't work either.

Another issue is that the database content is nested, which means I have subtrees, subsubtrees etc. For example, for the entry `first.name="John", last.name="Smith"` I might have subentries like `address, age, occupation`, and for occupation again I might have categories as subtrees (e.g. years from 2005 to 2012 and for each year I would have an an entry like "unemployed", "clerk", "entrepreneur"). So what if I want to find all people with first name "John" who are 40 years old and were unemployed in 2010? What would the query look like?

EDIT as a reply to Stennie: Here's an example of the structure of my database and the query I'm trying to do. Imagine that alumnis of a university have been subdivided into groups (e.g. "very good students", "good students" and so on). Each group then contains a list of people that have been assigned to this group along with their details.

```
(0){..}
   _id  : (Object ID) class id
   groupname: (string) unique name for this group (e.g. "beststudents")
   members[11]
       (0){..}
           persid : (integer) 1
           firstname: (string)
           surname: (string)
           age: (integer)
           occupation: (string)
       (1){..}
           persid : (integer) 2
           firstname: (string)
           surname: (string)
           age: (integer)
           occupation: (string)
#      and so on until (10){..}
(1){..}
```

```
    _id  : (Object ID) class id
    groupname: (string) unique name for this group
    members[3]
        (0){..}
            persid : (integer) 1
            firstname: (string)
            surname: (string)
            age: (integer)
            occupation: (string)
#      and so on until (2){..}
# and many more
```

Now let's assume that I am interested in the groups with the names "best students" and "good students", and would like to get "surname" and "occupation" for each member of each of these groups as an R object in order to do some plots, stats or whatever. And maybe I'd also want to refine this request to only get those members that are younger than 40 years old. Now after having read Stennie's reply, I tried it this way:

```
cursor <- mongo.find(mongo, "test.people",
          list(groupname=list('$in'=c("beststudents", "goodstudents")),
               members.age=list('$lt'=40) # I haven't tried this with my DB, so I hop
               ),
          fields=list(members.surname=1L, members.occupation=1L)
        )
count <- mongo.count(mongo, "test.people",
          list(groupname=list('$in'=c("beststudents", "goodstudents")),
               members.age=list('$lt'=40)
               )
        )
surnames <- vector("character", count)
occupations <- vector("character", count)
i <- 1
while (mongo.cursor.next(cursor)) {
  b <- mongo.cursor.value(cursor)
  surnames[i] <- mongo.bson.value(b, "members.surname")
  occupations[i] <- mongo.bson.value(b, "members.occupation")
  i <- i + 1
}
df <- as.data.frame(list(surnames=surnames, occupations=occupations))
```

There's no error message after running this, but I get an empty data frame. What's wrong with this code?

r mongodb rmongodb

edited Sep 13 '12 at 12:13
Stennie
18.8k 3 32 56

asked Aug 17 '12 at 9:42
AnjaM
594 6 19

Can you post an example document? Ideally a reduced example which shows the typical nesting for your query. – Stennie Aug 22 '12 at 13:09

add a comment

2 Answers

> Now, what if I want to find people whose first name is either "John" or "Bob" or "Catherine"?

You can use the `$in` operator for this:

```
cursor <- mongo.find(mongo, "test.people",
   list(last.name="Smith",
        first.name=list('$in'=c('John','Bob','Catherine'))
   )
)
```

It would be worth having a read of the MongoDB Advanced Queries page as well as Dot Notation (Reaching Into Objects).

> Another issue is that the database content is nested, which means I have subtrees, subsubtrees etc.

The data structure sounds potentially challenging to manipulate; would need a practical example of a document to try to illustrate the query.

> So what if I want to find all people with first name "John" who are 40 years old and were unemployed in 2010? What would the query look like?

Making some assumptions on the data structure, here is an example of a simple "and" query:

```
cursor <- mongo.find(mongo, "test.people",
    list(
        first.name='John',
        fy2012.job='unemployed',
        age = 40
    )
)
```

edited Aug 22 '12 at 13:50

answered Aug 22 '12 at 13:32
Stennie
18.8k 3 32 56

Thanks for your reply, it's helpful. I inserted a more detailed example with a code that I've tried out after having read your reply, but that didn't work the way it should. Do you have an idea about what's the problem? – AnjaM Aug 28 '12 at 15:20

@AnjaM: Apologies, missed your last comment here which also adds a new example in the description. Were you able to sort out the issue with that query? If not, it is probably better to add as a separate question. – Stennie Sep 13 '12 at 12:09

add a comment

It's not really an answer as I'm still struggling with some aspects myself, but this might help to get you started: Running advanced MongoDB queries in R with rmongodb

Also, check the example application that ships with the rmongodb page. That is, which is accessible at the github package: https://github.com/gerald-lindsly/rmongodb/blob/master/rmongodb/demo/teachers_aid.R

answered Aug 17 '12 at 12:18
Rappster
2,495 17 36

add a comment

Not the answer you're looking for? Browse other questions tagged r mongodb rmongodb or ask your own question.