



**Eötvös Loránd Tudományegyetem
Informatikai Kar**

Programozási nyelvek és fordítóprogramok tanszék

**Lekérdező nyelv definiálása és prototípus
implementálása kódmegértés céljából**

Témavezető:
Brunner Tibor
PhD hallgató

Készítette:
Bakos Péter
MSc hallgató

Budapest, 2017

Tartalomjegyzék

| | |
|---|-----------|
| 1. Bevezetés | 2 |
| 2. Lekérdező nyelv | 3 |
| 2.1. jQuery | 3 |
| 2.2. .QL | 4 |
| 2.2.1. Select utasítások | 4 |
| 2.2.2. Predikátumok | 4 |
| 2.2.3. Aggregátor függvények | 5 |
| 2.2.4. Osztályok | 5 |
| 2.2.5. Generikus lekérdezések | 5 |
| 2.3. Program Query Language | 6 |
| 2.4. Soul | 7 |
| 2.5. Java Tools Language | 8 |
| 2.5.1. Egyszerű minták | 8 |
| 2.5.2. Változók | 9 |
| 2.6. Összehasonlítás | 10 |
| 3. Compass Query Language | 12 |
| 3.1. Nyelv tervezés és fejlődés | 12 |
| 3.1.1. Függvény lekérdezés szignatúra szerint | 12 |
| 3.1.2. Változó olvasás/írás | 13 |
| 3.2. Szintaktika | 13 |
| 3.2.1. JavaScript Object Notation | 13 |
| 3.2.2. CQL szintaktika | 14 |
| 4. Konklúzió | 19 |

1. fejezet

Bevezetés

2. fejezet

Lekérdező nyelv

Lekérdező nyelvvel a szoftverfejlesztéssel foglalkozók már mind találkoztak: relációs adatbázisok adatainak kinyeréséhez, hozzáadásához, törléséhez és módosításához használjuk az SQL (Structured Query Language) nyelvet. Az SQL egy szakterület-specifikus nyelv (angolul domain-specific language, röviden DSL), amely egy bizonyos szakterületre koncentrál (adatbáziskezelés). Az SQL az adatbázis információinak lekérdezéséhez egy néhány utasításból álló parancsot használ. A nyelv komplexitását mutatja, hogy parancsokon belül is lehetnek alparancsok, ezáltal kifinomultabb lekérdezéseket is alkalmazhat a felhasználó. Ezzel ellentétben a LINQ (Language Integrated Query) egy olyan lekérdező nyelv, amely a C# programozási nyelv része. A LINQ-t a Microsoft fejlesztett ki C# nyelv adatstruktúrák információjának kinyerésére az SQL nyelv szintaktikájában használatos utasításokkal (Select, Where, Group By, stb.). A forráskódban való kereséshez az idők, ahogy az ipari szoftverek terjedelme egyre nőtt, úgy nőtték a forráskód sorainak száma is. Egy ipari méretű szoftverben a kiigazodás jelentősen komplex feladat. Az egyszerű szöveges keresők nem elég okosak ahhoz, hogy olyan kereséseket végezzenek. Ennek a műveletnek a megkönnyebbítésére az évek során több lekérdező nyelvet is fejlesztettek. Az alábbiakban összehasonlítom a piacon levő már kiforrott lekérdező nyelveket kifejező erő, komplexitás és a felhasználó tanulása szempontjából.

2.1. jQuery

A JQuery lekérdező nyelv a logikai programozási nyelvek közé sorolható, ami TyRuBa programnyelven alapul [4]. A TyRuBa hasonlít a Prologhoz, azaz egy olyan logikai programozási nyelv, amely rugalmasságot biztosít komplex lekérdezések írásához és szabályokat magasabb rendű kapcsolatok leírásához. A JQuery lekérdező nyelv valójában a TyRuBa nyelv kiterjesztése egy olyan könyvtárral, amikben a forráskódhoz kapcsolódó predikátumok vannak definiálva. Minden predikátumra a nevével hivatkozunk, a paraméterei lehetnek változók (amik elejére kérdőjelet teszünk) vagy attribútum nevek. Például a `class(?C, name, HelloWorld)` lekérdezés visszaadja az összes olyan osztályt, aminek a `name` tulajdonsága („property”-je), azaz az osztály neve a `HelloWorld` szöveggel megegyezik. Komplexebb lekérdezéseket úgy kapunk, ha vesszővel elválasztott lekérdezés-sorozatot írunk. Tegyük fel, hogy keressük az olyan osztályokat, amiknek a létezik olyan függvénye, aminek visszatérési értéke int. Ezt a lekérdezést így írhatjuk le:

```
class(?C, method, ?M),
```

```
method(?M, returnType, int)
```

2.2. .QL

A .QL egy olyan lekérdező nyelv, amely külsőre nagyon hasonlít az SQL nyelvre [2]. A szoftverfejlesztők által jól ismert SELECT-FROM-WHERE hármásra alapoznak, viszont a könnyebb olvashatóság érdekében a szelekció került a lekérdezés végére (FROM-WHERE-SELECT). A hasonlóság azonban kimerül a szintaktikában, a szemantikája teljesen eltérő az SQL nyelvtől. Így a .QL lekérdezések írása könnyen elsajátítható bármely szoftverfejlesztésben minimálisan is jártas felhasználó által az SQL szintaxisát alkalmazva. A .QL a Datalog nyelven alapszik, amely egy egyszerű logikai programozási nyelv. Mivel Datalogban rekurzív lekérdezések is használhatók, így a hívási gráf vagy az öröklődés hierarchiája is egyszerűen, könnyen alkalmazható. A .QL egy objektumorientált lekérdező nyelv, amelyben az osztályok predikátumok, az öröklődés pedig az implikációt jelenti. Ezáltal a lekérdezések újrahazsnosíthatók, felhasználhatók akár más projektekben is.

2.2.1. Select utasítások

A Java programozási nyelvben az objektumok egyenlőségvizsgálatára az equals() függvényt használjuk. Ha egy osztály definiál saját equals() függvényt, akkor annak implementálnia kell a hashCode() függvényét is. Ez amiatt szükséges, ha hash alapú adatszerkezetbe rakjuk az adott osztály objektumait, akkor az objektumok ekvivalenciájának vizsgálata ne sérüljön. A lekérdező nyelvek segítségével egyszerűbben felderíthetők az ilyen hibák a forráskódban: keressük meg az olyan osztályokat, amik deklarálják az equals() függvényt, de a hashCode()-ot nem. Ez a lekérdezés a .QL-ben így néz ki:

```
from Class c
where c.declaresMethod("equals") and
      not (c.declaresMethod("hashCode")) and
      c.fromSource
select c.getPackage(), c
```

A Select rész hasonlóan az SQL-hez, kilistázza ezen osztályok csomagnevét és magának az osztálynak a nevét.

2.2.2. Predikátumok

Az elnevezett és parametrizált lekérdezéseket nevezzük predikátumoknak. Ezeket a predikátumokat felhasználhatjuk más lekérdezésekben is. A predikátumok által egyszerűen kaphatunk rendkívül komplex lekérdezéseket anélkül, hogy feláldoznánk az olvashatóságát a lekérdezésnek, hiszen megfelelő névválasztással a predikátumok absztraktabbá, tömörebbé tehetik azokat.

```
predicate between(RefType down, RefType between, RefType up){
  down.hasSuperType*(between) and
  between.hasSuperType*(up)
}
```

A fenti predikátum használatával megkaphatók a *down* és *up* típusok öröklődési hierarchiájában a köztes osztályok. Azaz, a predikátum igaza lesz, ha a *down* osztály a *between-nek* altípusa (leszármazottja), és az *up* a *between-nek* szupertípusa (őse).

2.2.3. Aggregátor függvények

Az SQL nyelv lehetőséget ad különböző összegek, átlagok, számlálások kiszámítására. Ezek az aggregátor függvények használata nehézkes, gyakran csak egymásba ágyazott utasításokkal értelmezhető. A .QL az Eindhoven Kvantor Jelölést használja, ami megkönnyíti az előbb említett függvények írását. Ahogyan SQL-ben, úgy .QL-ben is a SELECT utasításhoz írjuk az aggregátor függvényt. Azonban a GROUP BY utasítás nem szükséges .QL nyelvben. Az aggregátor függvény három részből áll: lokális változók, feltételek, term. Ezek a '—' karakterrel vannak elhatárolva és ezeket előzi meg az aggregátor függvény neve (sum, count, avg, min, max). A lokális változókat felhasználhatjuk a feltételekben, amik értelem szerűen szűrik a találatokat. A term határozza meg, hogy mely értéken szeretnénk aggregálni. A sorrend és azok szerepe nagyban hasonlít a FROM-WHERE-SELECT hármásra.

```
from Package p
where p.hasName("abc.aspectj.ast")
select sum(CompilationUnit cu |
           cu.getPackage() = p |
           cu.getNumberOfLines())
```

Ezzel a lekérdezéssel és a *sum* aggregátor függvény használatával megkapható az *ast* csomagban található forráskód sorainak száma.

2.2.4. Osztályok

.QL nyelvben a predikátumok mellett meghatározhatunk komplexebb entitásokat: osztályokat. Osztályok segítségével definiálhatunk új típusokat a FROM szekcióhoz. Az osztály „konstruktor” megadja, hogy mit várhatunk a típustól: például milyen osztálynak a leszármazottja kell, hogy legyen, vagy milyen függvényeket kell tartalmaznia. Definiálhatunk az osztályhoz olyan függvényt, amellyel megadhatjuk a keresni kívánt típusnak egy adattagját, annak típusára szintén tehetünk megszorítást.

```
class HasIntField {
    Field getIntField() {
        result = this.getAField() and
        result.getType() = int
    }
}
```

Ezzel az osztállyal szűrhetők az *int* típusú adattagokat tartalmazó osztályok. Ezt az osztályt alkalmazhatjuk bármely lekérdezésben.

2.2.5. Generikus lekérdezések

A .QL objektumorientált megközelítésének az előnyeihez sorolható, hogy megírt lekérdezések alkalmazhatók több projektben is. A lekérdezéseket még általánosabbá

teszik a generikus lekérdezések, amiket gyorsan és könnyedén felhasználhatóvá válnak bármely kódázishoz. Vegyük például a szoftverfejlesztésben gyakran alkalmazott tervezési mintákat (design pattern). A .QL nyelvben lehetőség van a Gyár tervezési minta leírására. A Gyár egy olyan speciális osztály, amely több típusú, de valami által összeköttetésben levő objektumokat állít elő, ezzel nagyobb rugalmasságot adva az objektumok létrehozásához. Mivel a Gyár minta nem konkrét típusokkal írja le működését, így a lekérdezés sem tartalmaz explicite típusokat. De mivel a .QL objektumorientált, ezért egy lekérdezés öröközhet és újra is definiálhat függvényeket, amikkel újrafelhasználhatóság is növekszik. Így már konkrét Gyár osztályokat kereshetünk a lekérdezéssel.

2.3. Program Query Language

A PQL (Program Query Language) kissé eltér a már említett lekérdező nyelvektől, abban hogy csak és kizárólag a forráskódban fellelhető minták keresésére alkalmas [5]. Így egyszerűbben megtalálhatók programozási hibák, figyelmetlenségek, biztonsági rések, memória szivárgások. Ilyen például nem lezárt File típusú objektumok keresése, vagy akár egy adatbázis könyvtár használatakor SQL injekció támadás elleni hibák felismerése. A lekérdezés primitív események szekvenciája. Ezen események objektumokon meghívott függvényekként lehet leírni, ahogyan a Java programozási nyelv szintaxisában megszokhattuk. Rekurzív lekérdezések írására is alkalmas a PQL. Ezt elnevezett allekérdezések segítségével lehet elérni. A felhasználó definiálhat különböző típusú változókat, amiket a lekérdezésben felhasználhat. A keresésben egy változó egy objektumhoz tartozik, a deklarációs részben megadott típusának vagy annak leszármazottjának kell, hogy megfeleljen az adott objektum.

```
query simpleSQLInjection()
uses
    object HttpServletRequest r;
    object Connection c;
    object String p;
matches {
    p = r.getParameter(_);
}
replace c.execute(p)
with Util.CheckedSQL(c, p);
```

A változók reprezentálhatnak adattagokat vagy tagfüggvényeket is, amik szövegesen kell, hogy megegyezzenek a mintában. A „*” karakter helyettesít bármely hosszú karakterláncot. Ha egy adattagot talál a mintában, akkor annak előfordulása a mintában mind ugyanahhoz az objektumhoz kell tartoznia. Lehetőség nyílik a „-” helyettesítő karaktert használni, ami a „*” karaktertől eltérően különböző objektumhoz vagy adattaghoz tartozik. Azonban, az így megtalált szimbólumokat nem lehet vizsgálni vagy visszatéríteni. A PQL lekérdező nyelv változói lehetnek argumentumok (amit egy másik lekérdezésből a jelenlegi lekérdezést meghívva kapott paraméterül), visszatérési érték (a hívó lekérdezés számára visszatérítendő változó), vagy belső változó (amik csak a lekérdezésben használhatók lokálisan).

Az egyszerű parancsokat (függvényhívások) össze lehet kombinálni összetett parancsokká. Az a;b sorozat az „a” megelőzi „b”-t jelöli. Általában, ez azt is jelen-

ti, hogy a két esemény között több esemény is előfordul – a lényeg a két egyedi eseményen alapszik, tehát a szekvenciák általában nem folytonosak. A nyelv lehetőséget kínál arra, hogy egy esemény be nem következését is lehessen definiálni: az `a; b;c` jelölésben „a” megelőzi „c”-t, úgy hogy a „b” esemény a kettő között nem szerepel. Szintén, lehetőség nyílik az alternáló operátor segítségével („—” karakter) több esemény közül csak az egyik bekövetkezését meghatározni. A `within` kulcsszó után adható meg, hogy a mintakeresés melyik függvényben történjen. A Java nyelvben a fájlkezeléshez szükséges objektumokat manuálisan kell felszabadítani, különben a rendszer nem kapja vissza az erőforrást a programtól.

```
query forceClose()
uses object InputStream in;
within _._();
matches {
    in = new InputStream();
    ~in.close();
}
executes in.close();
```

A fenti lekérdezés megtalálja az összes olyan függvényt, amelyben a megnyitott fájlokat nem zárja be a programozó. A lekérdezés akkor ad találatot, ha van egy olyan függvény, amiben van egy `InputStream` objektum viszont a függvény végéig bezárólag nincs meghívva az objektumon a `close()` tagfüggvény. Mivel nem csak az adott függvény zárhatja le az `InputStream`-et, hanem az abból hívott függvények is, ezért a `within close()`-ban a „_” helyettesítő karakter minden függvényhívást is megvizsgál.

A PQL lekérdezések gyakran átvizsgálandó vagy nem kívánt program viselkedést határoznak meg. PQL-ben két lehetőség nyílik az információ logolásra vagy javítási akció végrehajtására. A legegyszerűbb eset az `executes` klóz. Az ebben leírt függvény akkor fog meghívódni, amikor a keresés találatot kap. A másik mód amikor rögtön javítani szeretnénk a kódot, azáltal hogy kódot helyettesítünk, már meglévő helyére. Ez a `replaces` részben tehető meg: a `replaces` kulcsszó utáni kifejezés helyett lesz beillesztve az ezt követő `with` kulcsszó utáni kódrészlet.

2.4. Soul

A SOUL egy logikai-alapú program lekérdező nyelv [3]. A felhasználó megadja a program karakterisztikáját, struktúráját logikai feltételek segítségével, majd ezt a SOUL már egy konkrét programkód mintájára illeszti. Minden lekérdezés az `if` kulcsszóval kezdődik és az abban található összes logikai változó első karaktere a kérdőjel. Mint minden logikai nyelvben, itt is predikátumokkal lehet leírni a nyelvben, amit a felhasználó szeretne. A predikátumok előre definiáltak, használatukkal lehet felépíteni egy lekérdezést. Vegyünk egy egyszerű példát: határozzuk meg az olyan függvénypárokat ahol az első valamely mélységben meghívja az utóbbit.

```
if ?outer isStatement,
    ?inner isStatementIn: ?outer
```

A lekérdezés két feltételt tartalmaz. Az első egy unáris predikátum, ami az `?outer` változót köti az egyik függvényhez a forráskódban az `isStatement` predikátum

segítségével. A második feltétele a lekérdezésnek egy bináris predikátum, ami a már előzőleg kötött `?outer` változót használja fel, és bevezet egy `?inner` változót, amit az `isStatementIn` predikátum fog kötni egy olyan függvényhez, amit az `?outer` változó által kötött függvényen belül hívtak meg. A predikátumokat vesszővel választjuk el egymástól. A forráskód alakjához, struktúrájához, felépítéséhez sokkal közelebb álló kód sablonokat is lehet definiálni a SOUL lekérdezésekben. A kód sablon egy funktort tartalmaz, amit egy paraméter követ. Ez után kapcsos zárójelek között a kódrészlet. A sablon funktora definiálja, hogy mely nyelvi elemek között kell keresni a lekérdezés, majd ezek után a funktoron belüli kódrészlet megtalálása a feladat. A kódrészlet egyfajta keverése a Java és a logikai programok szintaxisának.

```
if jtClassDeclaration(?classDeclaration){
    class ?className{
        private ?fieldDeclarationType ?fieldName;
        ?modifierList ?returnType ?methodName(?parameterList){
            return ?fieldName;
        }
    }
}
```

A fenti lekérdezés az olyan osztályokat keresi meg, ahol az adatagok el vannak rejtve az osztály használóitól és csak getter függvény hívásokkal érhetők el.

2.5. Java Tools Language

Ahogy azt a neve is mutatja a Java Tools Language (JTL) egy Java forráskód keresést megkönnyebítő lekérdező nyelv [1]. A JTL szintaxisa tömör és intuitív bármely Java nyelvben gyakorlott programozónak. Egy JTL lekérdezés sok esetben pontosan megegyezik egy Java osztály- vagy függvénydefinícióval. Ez növeli a lekérdezés olvashatóságát és nem utolsósorban a felhasználó magától értetődően tudja JTL mintákat írni. A nyelv két fő adattípusa a `Member` és a `Type`. A `Member` fogalmába tartoznak az osztályok és interface-ek tagjai, beleértve a tagfüggvényeket, adattagokat, konstruktorokat, inicializálókat. A `Type` adattípusba a Java osztályokat, interface-eket, enum-okat és a Java nyelv primitív típusait (pl. `int`, `float`) értjük. Egy JTL program nevesített logikai predikátumok összessége. Egyet kiválasztva a végrehajtott predikátum lesz egy lekérdezés. A predikátumnak kis betűvel kell kezdődjön, a változóknak pedig nagyval.

2.5.1. Egyszerű minták

A Java nyelv kulcsszavai visszaköszönnek a JTL nyelvben, amik ugyanazt a szemantikai jelentést hordozzák, mint amit a forráskódban a Java fordító is értelmez. Például, az `int` kulcsszó a Java nyelvben állhat egy függvény előtt, mint visszatérési érték, vagy adattag előtt, mint az annak típusát jelezve. A JTL nyelvben ugyanezek a szemantikai jelentésük van. A `public` minta minden olyan program részletet megtalál, ami publikus láthatóságú osztály vagy például publikus elérési tagfüggvény, mező. Azonban, vannak a JTL-hez hozzáadott a Java nyelvben nem kulcsszóként ismert natív kulcsszavai. Például az `anonymous` `Type`-okra van értelmezve, amely

névtelen osztályokat keres. A JTL-nek kétféle predikátuma létezik: natív és összetett. A natív predikátumok implementációját nem tartalmazza a nyelv, azaz ha szeretnénk egy ilyen predikátumot kiértékelni, akkor azt csak külső könyvtárak segítségével tehetjük meg. Ezzel ellentétben, az összetett predikátumok definiáltak egy logikai operátorokat használó JTL kifejezés által. A vesszővel elválasztott kulcsszavak a konjugációt jelölik.

```
class {  
    public int method;  
}
```

Tehát, a `public`, `int` mintára az összes olyan publikus függvény fog illeszkedni, aminek a visszatérési értéke `int` típusú (illetve természetesen az `int` típusú publikus adattagok is a keresés eredményébe beletartoznak). A vessző elválasztó azonban csak opcionális, elhagyása szintén konjugációt jelent. A diszjunkciót a függőleges vonallal (`|`) jelezhetjük, míg a negációt a felkiáltó jellel. Következésképpen, a `!public` ekvivalens kifejezés a `protected` | `private` kifejezéssel. Az újrafelhasználást szem előtt tartva, ezeket a mintákat elnevezhetjük, így már csak a definiált kereső kifejezést használhatjuk. Reguláris kifejezések alkalmazása szintén támogatott a JTL nyelvben. Függvények, osztályok, adattagok nevei helyett lehet használni a reguláris kifejezéseket. A függvények paraméterlistájában lehetőség nyílik helyettesítő karakterek használatára. A `'*` karakter nulla vagy több paraméternek felel meg, míg a `'_'` karakter pontosan egy típust helyettesít.

2.5.2. Változók

A változók szerepe a nyelvben hasonlít a Prolog logikai programozási nyelvben használtakhoz. Például ha egy változót kétszer is használunk egy lekérdezésben, akkor annak ugyanazt a típusnevet, függvénynevet vagy osztálynevet kell jelentenie. Az alábbi lekérdezésnek olyan függvények felelnek meg, amiknek a visszatérési értékének típusa megegyezik az argumentumlistájában szereplő valamely paraméter típusával.

```
return_arg := RetType (*, RetType, *);
```

A változók használatának másik területe a predikátumokhoz köthető. A JTL több saját predikátumot definiál, aminek nagy része Java nyelvbeli kulcsszavak. Ilyen az `implements[I]` predikátum, ami egy `Interface` típust vár, és visszaadja az összes osztályt, ami implementálja azt. Minden predikátum tartalmaz egy rejtett paramétert, a „vevője” a mintának, amit a `This` vagy `'#'` karakterrel lehet hivatkozni. A natív JTL predikátumok közé tartozik többek között: `members[M]` (igaz, ha `M` `This`-nek adattagja, akár definiált, akár örökölt), `overriding[M]` (igaz, ha `This` egy olyan függvény, ami felüldefiniálja `M`-et).

```
container[C] := C.members[This]
```

A `container` egy olyan JTL által alapértelmezett egyváltozós predikátum, ami akkor teljesül, ha egy osztálynak van a `C` paraméterű adattagja.

A programozó írhat saját predikátumot is különböző paraméterekkel. Ahogy az a logikai programozásban is lenni szokott, a predikátumok paraméterei nem mások, mint külsőleg elérhető változók. Az alábbi predikátum például egy paramétert fogad.

Amikor egy bizonyos értékkel lesz meghívva, akkor az argumentumnak megfelelő típusú statikus adattagok fognak megfelelni a keresésnek.

```
copty_ctor := constructor(T), T.members[This]
```

A C++ nyelvben egy objektum másolását a másoló konstruktor (copy constructor) végzi. A JTL által ezt a fentebbi lekérdezés szerint írható le. Tehát, egy osztálynak akkor van másoló konstruktora, ha létezik olyan konstruktor, amely egy a ugyanazon típusú objektumot vár paraméterként.

2.6. Összehasonlítás

A vizsgált lekérdező nyelvek közül három logikai programozáson alapul (jQuery, SOUL, JTL), kettő objektum-orientált megközelítést alkalmaz (PQL, .QL). A kódmegértés oka lehet forrás kód refaktorálás. Ez az a folyamat, amikor úgy változtatjuk meg a már meglévő forráskódot, új struktúrát adva, hogy közben a program viselkedése a külső szemlélő számára nem változik meg. Refaktorálás általánosan két esetben szokás alkalmazni: új funkció bevezetésénél (ha a jelenlegi kód nem alkalmas az új implementálásra, vagy megkönnyíti azt), illetve ha bug javítása ezt kívánja meg. Az utóbbi művelethez csak a PQL nyújt lehetőséget, ami szintén a programozó munkáját segíti, hiszen a nyelv alkalmazásával automatikusan lehet az adott hibákat felkutatni és kijavítani.

Vizsgáljuk meg a lekérdező nyelveket a mindennapi kód kereséshez: találjuk meg az összes *f* nevű függvényt. Majd elemezzük a lekérdezés olvashatóság, kifejező erő, újrafelhasználhatóság szempontjából. jQuery nyelvben ez a lekérdezés az alábbiak szerint írható le:

```
method(?M, name, f)
```

Egyszerű, könnyen olvasható még a jQuery nyelvet nem ismerő számára is. .QL-ben ugyanez a kifejezés az SQL szintaxisnak köszönhetően ugyancsak rendkívül jól olvasható.

```
From Method m
Where m.hasName("f")
Select m
```

A SOUL lekérdező nyelvben mivel egy függvény struktúráját kell leírni az olvashatóság szempontjából az ilyen egyszerű lekérdezésekben is elveszhet a felhasználó.

```
if jtMethodDeclaration(?m){
  public ?type f(?parameterList){
    ?statements
  }
}
```

A JTL keressük meg az összes *f* nevű függvény lekérdezése talán a legtisztább mind közül. Pontosan tükrözi, mit keres a programozó, és azt röviden, tömören lehet leírni.

```
public f(*)
```

A PQL nyelv fejlesztésekor az alapvető programozási hibák és bugok feltárása volt a cél. Ezt úgy érték el, hogy program végrehajtási gráfján lehet keresni, így csak függvényhívás helye határozható meg. Következésképpen, PQL-lel nem lehet ki-listázni a forráskódban fellelhető f nevű függvényeket, csak annak hívásait.

3. fejezet

Compass Query Language

3.1. Nyelv tervezés és fejlődés

A lekérdező nyelv tervezésekor elsődleges célok között voltak a nyelvfüggetlenség, könnyű olvashatóság és a magas kifejező erő. A nyelv fejlesztése a való életből vett példák leírásával történt. Ezek olyan lekérdezések, amelyek egy szoftver fejlesztőnek rögtön eszébe jutna, ha új projektre, modul fejlesztésére kerül. Ezen felül olyan általános és összetett vagy komplex lekérdezések kerültek a példákba, amik segítik a lekérdező nyelv definiálását: például allekérdezések, attribútumok meghatározása. A tíz példa kérdés használatával már körvonalazódik a nyelv. Egyesével vizsgálva őket, de egy bizonyos egységet megtartva a lekérdezések között, néhány meggon-
dolással a kialakítható a nyelv.

3.1.1. Függvény lekérdezés szignatúra szerint

Lekérdezés: Határozzuk meg az összes *void* visszatérési értékű függvényeket, amelyek második paramétere *int* típusú.

A legalapvetőbb kérdések egyike függvények megkeresése annak szignatúrája alapján. Egy függvény szignatúrájának három része van: visszatérési érték, név, paraméterlista. Az előbbi kettő egyértelműen egyszerű értékek, karakterláncok. A paraméterlista, ahogy a neve is mutatja egy listának kell lenni.

Egy lista megkülönböztetésére egy egyszerű szövegtől a négyzetes zárójelek alkalmazandók. A listában való elemek elkülönítésére a vessző karaktert kell használni. Elképzelhető, hogy a lista nem minden elemét ismeri a programozó, vagy egyszerűen csak nem releváns információval bír az adott paraméter. Az ilyen problémák megoldására lehet használni reguláris kifejezéseket. Fontos, hogy könnyen lehessen csak egy vagy több argumentumot helyettesíteni. Ennek érdekében két helyettesítő karakter kerül bevezetésre: '_' és a '?' karakterek. Az előbbi egy argumentum kiváltására szolgál, míg utóbbival nulla vagy több paramétert van lehetőség pótolni. Az akárhány karakter helyettesítésére általában a '*' karaktert szokták definiálni. A nyelv esetében azonban problémás lehet, ha C/C++ nyelveket vesszük figyelembe, hiszen ezekben a programozási nyelvekben a mutatók típusát a csillag karakter jelöli. Annak érdekében, hogy a lekérdező nyelv független maradjon a programozási nyelvektől, a kérdőjelre esett a választás, mivel a kérdőjel nem szerepel típus nevében egyetlen széles körben használt nyelvben sem.

A reguláris kifejezés használatával már egyszerűen megadható a példában kere-

sendő függvények definíciója. A lekérdezésnek három része van: függvényt keresünk, visszatérési értéke *void*, második paramétere *int*.

```
function {  
    return: void  
    arguments: [_, int, ?]  
}
```

Először jelezzük, hogy a függvényeket vizsgáljuk. Visszatérési értéknek (*return*:) megadjuk a *void* szót. A második paraméter legyen *int* típusú kifejezéssel ekvivalens az első paramétert nem ismerjük, második legyen *int*, és utána következhet akármennyi (akár nulla) argumentum is. Ezzel az ekvivalens állítással kapjuk a fenti paraméterlistát (*arguments*:).

3.1.2. Változó olvasás/írás

Lekérdezés: Határozzuk meg egy változó írását/olvasását.

3.2. Szintaktika

Felhasználói szempontból egy új nyelv megtanulása mindig visszatartó erő lehet a nyelv elsajátítására. Ezért egy olyan lekérdező nyelv megalkotása volt a célom, amelyet könnyű elsajátítani, megtanulni, intuitív kezelni és nyelvfüggetlen. Az előző fejezetben vizsgált lekérdező nyelvek közül volt olyan, ami ezt úgy kívánta elérni, hogy minden programozó által ismert nyelv szintaktikájára alapoz (SQL). A Java Tools Language a Java nyelvhez igazította a szintaktikáját, megkönnyítve ezzel az abban dolgozókat. Azonban, az így megalkotott nyelv elveszti nyelvfüggetlenségét. A két szempontot szem előtt tartva (létező nyelvre alapozni és nyelvfüggetlen maradni) a JavaScript nyelvhez kifejlesztett, de azóta több nyelvben is alkalmazott JavaScript Object Notation.

3.2.1. JavaScript Object Notation

A JavaScript Object Notation (JSON) a nevében szereplő scriptnyelvből, a JavaScript-ből alakult ki [6]. A JSON egy olyan objektum, ami az XML-hez képest kis méretű, adatcserére alkalmazott formátum. Az egyszerű formátumának köszönhetően programok számára könnyen parszolható és generálható és mindeközben az ember számára is megmarad olvashatóság. Ezek mellett a JSON egy szöveg alapú objektum, így programozási nyelvtől is független.

A JSON objektumokat használata kiterjed kommunikációs protokollok implementálására és konfigurációs információk reprezentálására. Tulajdonképpen, bármilyen információ küldésére, fogadására alkalmas formátum.

A JSON struktúrája név/érték párokból tevődik, amit egyszerű módon lehet implementálni bármely programozási nyelv által. Tehát, a JSON egy sorrendtől független név/érték párok halmaza. Az objektum egy nyitó kapcsos zárójellel kezdődik ({}) és egy csukóval végződik (}). Egy név után kettőspont (:) választja el az értéktől, a párokat pedig vesszővel választja (,) el egymástól. Az értékek típusosak (ellenben az XML-lel), amik lehetnek: karakterlánc, szám, tömb, logikai érték. A tömböket

négyzetes zárójelek ([]) között alkotott elemeket értjük, vesszővel elválasztva. Az elemek bármely JSON objektum lehet. Ezeken felül lehetnek összetett típusok, azaz az előbbi egyszerű típusok által alkotott összetett objektumok, illetve a *null* kulcsszóval jelölt üres objektum. A következő példában telefonkönyvben található bejegyzés egy lehetséges struktúrája látható.

```
{
  "firstName": "Janos",
  "lastName" : "Szabo",
  "address" : {
    "street" : "Kossuth Lajos u. 1.",
    "city" : "Budapest",
    "postal" : "1000"
  },
  "phoneNumbers" : [
    {
      "type" : "home",
      "number" : "06123456789"
    },
    {
      "type" : "mobile",
      "number" : "06987654321"
    }
  ]
}
```

A "firstName" és "lastName" egyszerű karakterláncok. Az "address" nevű objektum egy összetett típus, aminek van utca, város, irányítószám mezeje. A "phoneNumbers" mező egy tömböt reprezentál, ami összetett típusú objektumokat tartalmaz. Már ebből az egyszerű példából is látszódik, hogy mennyire összetett struktúrájú objektumok létrehozására alkalmas ez a formátum, és nem megy az olvashatóság rovására sem.

3.2.2. CQL szintaktika

A lekérdező nyelv szintaktikáját a JSON formátumra alapoztam. A választás teljesen megfelel a szintaktikával tett feltételeknek: a programozó számára legyen könnyen olvasható (a struktúra adja ezt a feltételt), elsajátítható (a JSON széles körben elterjedt már nem csak a JavaScript fejlesztők körében is). Azonban, csak a CQL alapját adja a JSON szintaktikája. Számos ponton ki kellett bővíteni a lekérdező nyelvet, például predikátumok bevezetésével, nevesített objektumokkal.

Entitások

A JSON objektumok névtelenek, csak a struktúra alapján lehet megmondani a típusukat. Egy lekérdező nyelvben viszont meg kell tudni mondani, hogy a programozó mire szeretne keresni (függvény, osztály, változó, stb.). Ezért CQL nyelvben entitásokat tartalmaz, amivel explicit meg lehet adni, hogy mely forráskódbeli

egységekben történjen a keresés. Az entitások meghatározásánál az objektum orientált paradigma részei játszottak szerepet. Az objektum orientáltság fő komponense az osztály. Az osztályt példányosítva kapjuk meg az objektumokat. Az osztály megadja, hogy az abból példányosított objektumoknak milyen szerkezetük van (adattagok), és milyen műveletek (tagfüggvények) végezhetők el rajtuk. Következésképpen, az osztálynak szerepelnie kell a lekérdező nyelvben. Természetesen az osztályt képező elemeket is tartalmaznia kell a nyelvnek, hiszen a kereséshez szükséges szűrőfeltételek ezek alapján is megadhatóak. Tehát, (tag)függvények és az adattagok is részét képezik az entitásoknak. Az adattagok lehetnek más objektumok is (nem csak primitív típusok, mint például az *int* vagy *bool*), ezért érdemes a változókkal együtt kezelni. Az objektum orientált paradigmától eltávolodva egy program viselkedését az absztrakt szintaxis fa írja le. Ennek elemeit szintén lekérdezhetővé lehet tenni, amely elemek különböző utasítások lehetnek.

Minden entitás attribútumokkal rendelkezik. A JSON formátumhoz hasonlóan az attribútum egy név és érték párból tevődik össze. Minden entitás definiál különböző attribútumokat, amik szűrőfeltételeknek felelnek meg. Egy entitás a nevével kezdődik, majd kapcsos zárójelek között szerepelnek az attribútumai. Az attribútumok megadása opcionális. Tételezzük fel, hogy olyan függvényeket keresünk, aminek *int* típusú visszatérési értékkel rendelkezik, de a függvény nevét nem tudjuk. Ekkor, mivel a név attribútum nem kötelező, ezért arra nem is teszünk megszorítást, annak kiírása elhagyható. Ahogy a JSON formátum is megengedi az összetett objektumok beágyazását, úgy a CQL-ben is lehetőség van olyan attribútumokat definiálni, melyek értékének típusa lehet például osztály vagy változó. Erre lehet tekinteni úgy is, mint egy allekérdezés.

Változó Egy lekérdezésben a változó entitást a *variable* kulcsszóval kezdődik. Jelölhet például egy egyszerű változót egy függvényen belül, vagy akár lehet adattagja is egy osztálynak, objektumnak. Az attribútumok között egyértelműen szerepel a név és a névtér. Egy változó meghatározó jellemzője a típusa, ezért szűrő feltételt erre is lehet meghatározni. Ha egy osztály adattagjáról beszélünk, akkor azt a *class* attribútum segítségével adható meg. Itt (is) jön előtérbe, hogy az attribútumok értéke lehet egy másik entitás (azaz egy összetett objektum).

| Név | Típus | Leírás |
|-----------|---------|---|
| name | String | Változó neve. |
| namespace | String | Változó melyik névtérben található. |
| type | String | Változó típusa. |
| class | Osztály | A változó melyik osztálynak az adattagja. |

Függvény A másik entitás, amik az osztály szerves részét képezik, azok a tagfüggvények. Mivel egy program függvények hívásából áll, ezért ennek a komponensnek a vizsgálata elsőrangú szerepet játszott a lekérdező nyelv megalkotásakor. Egy függvény neve mellett a szignatúrája az, ami meghatározza. Ennek megfelelően egy keresés szűrőfeltétele lehet a visszatérési érték típusa vagy a paraméterlistája. A paraméterlista jól illeszkedik a JSON formátumban ismert tömb típusra. A típusokat a nyitó- és zárójelek közé tesszük ('[', ']'). Előállhat olyan eset, hogy valamely paraméterre nem vagyunk kíváncsiak, vagy csak nem ismerjük. Az ilyen problémák megoldására használatos speciális karakterek, reguláris kifejezésekhez. Az egy paramétert kiváltó

karakter a `'_'`, a többet `'?'`. Az általánosan használt `''` karakter a C/C++ programozási nyelvekben a pointer-t jelenti, ezért esett a választás a kérdőjelre. Így olyan kifejezések is kereshetők, mint például a második paramétere a függvénynek legyen *int*. A paraméterlistához tartozik még, annak számossága, amire szintén lehet hivatkozni az *argumentCount* attribútummal.

A változóhoz hasonlóan, a függvénynek is megadható, hogy mely osztály tagját képezi. Ehhez kapcsolódóan az elérési szintjét is meg lehet adni: *private*, *protected* vagy *public*.

Egy programban való eligazodást nagyban nehezíti a függvények hívása, ami hosszú hívási láncot is eredményezhet. Tehát, fontos, hogy függvényen belül milyen más eljárások hívódnak. Ennek az inverze is rendkívüli fontossággal bír: egy függvényt a forráskódban hol hívnak. Mindkét attribútum értéke egyben allekérdezés is egy függvényre, így a programozónak még több lehetősége van testre szabni a keresett eljárást.

| Név | Típus | Leírás |
|---------------|----------|---|
| name | String | Függvény neve. |
| namespace | String | Függvény melyik névtérben található. |
| class | Osztály | A változó melyik osztálynak az adattagja. |
| return | String | Visszatérési érték típusa. |
| arguments | Lista | A függvény paraméterlistája. |
| argumentCount | Szám | A függvény paramétereinek száma. |
| access | String | Tagfüggvény esetén annak elérési szintje. |
| call | Függvény | A függvény mely függvényt hívja. |
| callFrom | Függvény | A függvény mely függvényből hívódik. |

Osztály Az osztályok különböznek az eddig említett entitásoktól. Egy osztályt reprezentáló objektum a CQL-ben nem csak attribútumokat tartalmazhat, hanem tagfüggvényeket és adattagokat. A JSON szintaktikája természetesen lehetővé tenné, hogy ezek szintén attribútumokként legyenek kezelve. A megvalósítást tömbök alkalmazásával lehet elérni. Azonban az olvashatóság megőrzése és az egyszerűség érdekében az osztályokba be lehet ágyazni függvény és változó lekérdezéseket. A lekérdezésből egyből látszik, hogy egy osztályon belül definiált függvények vagy változók az osztály részei. Tehát az attribútumok használata feleslegessé válna. Adattagokból és tagfüggvényekből természetesen bármennyit lehet definiálni (vagy akár egyet sem). Minden függvény vagy változó az osztálynak egy tagjára tesz megszorítást. Tehát, ha a felhasználó olyan osztályokat keres, aminek van *f* nevű függvénye, akkor az lekérdező osztály objektumon belül definiál egy ennek megfelelő függvény lekérdezést.

Az alapvető név és névtér attribútumokon kívül egy osztálynak lehet *inherit* tulajdonsága. Objektum orientált programozásban az újrafelhasználhatóság használatának egyik eszköze az öröklődés. Emiatt is fontos, hogy olyan attribútum is legyen, amivel szűrőfeltételnek megadható az őse. Magától értetődően ez az attribútum egy osztály allekérdezés, ahogyan a függvény lekérdezésnél a hívási tulajdonság, úgy itt is a testre szabhatóság miatt nem csak névre lehet szűrni, hanem egyéb attribútumokra is.

Az öröklődéshez kapcsolódik, de használatát tekintve elkülönítjük. Ez az interface osztályból való származtatás. Az interface (vagy absztrakt osztály) olyan

osztály, amit nem lehet példányosítani. Általában egy felületet biztosít és tagfüggvényeit csak deklarálja, nem definiálja. Ha egy osztály őse egy interface és annak összes absztrakt tagfüggvényét definiálja, akkor az osztály implementálja az interface-t. Ennek a speciális öröklődés is lehet kereséskor egy feltétel a felhasználó részéről, amit szintén egy osztályra való allekkérdezéssel lehet megoldani.

Az osztályok újrafelhasználhatóságának egy másik formája a generikus programozás. Lényege, hogy az osztályokat felparaméterezhetjük típusokkal, így még általánosítva a létrehozható objektumokat. Habár napjainkban a programozási nyelvek segítségével lehetőség van megszorításokat tenni a típusparaméterre, például C++ nyelvben ez csak a C++17-es szabványtól alkalmazható (concept-ek). A lekérdező nyelv által megadható a *template* attribútum, amivel megszorítás tehető a típusparaméterre. Ezzel olyan specializált osztályok kereshetők, amivel már konkrét típussal van paraméterezve, ezért példányosítható is.

| Név | Típus | Leírás |
|-----------|---------|---|
| name | String | Osztály neve. |
| namespace | String | Osztály melyik névtérben található. |
| inherit | Osztály | Milyen osztályból öröklődik. |
| implement | Osztály | Melyik osztályt valósítja meg. |
| template | Osztály | Az típusparaméterrel rendelkező osztály paramétere. |

Utasítás Az előző három entitástól teljesen különböző az utasítás. Egy program viselkedésének vizsgálatát legjobban az utasításainak megfigyelésével lehet megtenni. Ehhez szükséges a program absztrakt szintaxis fájának, AST-jének (Abstract Syntax Tree) a megléte. Ezzel megkapjuk a program összes utasítását, azok hívásait, milyen típusokkal, változókkal, függvényekkel működtek együtt, vagy hivatkoztak rájuk. Utasítást nem nevesítünk, így névvel nem is lehet hivatkozni egy ilyen elemre (nem is lenne értelme). Ezzel szemben viszont a típusára rákereshetünk: például ciklusra vagy elágazásra keresünk. Persze lehetséges ennél speciálisabb feltételt is megadni, mint például csak olyan *for* ciklus érdekel, amiben implicit konverzió történik az inicializáló részben definiált ciklusváltozóhoz. Mivel az utasítás nagyon általános megközelítés (ami lehet elágazás vagy visszatérés egy értékkel), nem lehet közös attribútumokat kitalálni rá. Ezért egyetlen tulajdonság, amik az attribútumok között szerepel az a típus (pl. switch, for, return).

| Név | Típus | Leírás |
|------|--------|----------------------------------|
| type | String | Az utasítást reprezentáló típus. |

Predikátumok

Az olvashatóság és újrafelhasználhatóság érdekében a nyelvnek tartalmaznia kell nevesített allekkérdezéseket. Ezen felül léteznek olyan tulajdonságok, amiket nem lehet kifejezni a CQL-ben levő attribútumokkal. Tegyük fel, hogy a forráskódban olyan osztályokat szeretnénk megkeresni, amiknek nincs egy bizonyos szignatúrájú függvénye. A lekérdezések eddig csak létezésre voltak alkalmazhatók, de jogos elvárás a felhasználtól, hogy ennek az ellentettjét is definiálni lehessen. A nyelv ezen részeit nevezzük predikátumnak. Tehát, a predikátumnak két típusa lehet: allekkérdezés vagy natív (attribútumokkal nem definiálható) predikátum. A lekérdező

nyelvben a predikátum eljárásként funkcionál. Következésképpen, paraméterek adhatók meg nekik. A formális argumentumok típus nélküliek, egyszerű karaktersorozatok. Az aktuális paraméter hasonlóan csak karakterlánc.

Minden predikátum a '@' jellel kezdődik, utána következik a neve. A '@' karakter emberi szemmel is jól elkülöníti az attribútumoktól.

Az allekérdezések szintaktikája a már említett '@' jellel kezdődnek, ezután következik a név, majd a formális argumentumlista gömböjű zárójelek között ('(', ')'). Ez megadja a predikátum szignatúráját. A szignatúra után a "==" (legyen operátor) következik, amit egy lekérdezés követ. Ez utóbbi lesz a predikátum definíciója. A gyakran használatos kifejezésekre minden programozási nyelvnek létezik sztemderd könyvtára. Ezek az általános fogalmak, eljárások a CQL-ben is megtalálhatók. Egy ilyen példa erre a C++ nyelvből ismert másoló konstruktor. A másoló konstruktor egy olyan konstruktor, ami egy másik ugyanolyan típusú objektumot vár paraméterül, mint amilyen objektumot konstruálni szeretnénk. Azaz, ahogy azt a neve is mutatja, egy meglévő objektumnak a másolatát hozzuk létre.

```
//Másoló konstruktor predikátum definíció
copy_ctor(T) := function {
  class: class { name: T }
  name: T
  arguments: [const T&]
}
```

A fentebb látható másoló konstruktor predikátum egy olyan függvény lekérdezés, aminek egy paramétere van (a T típus, aminek a másoló konstruktorát akarjuk megkapni), és három attribútumára lett téve megszorítás. Az attribútumok között megtalálható a *class*, amivel leszűrjük azokat a függvényeket, amelyek egy olyan osztálynak a tagfüggvénye, aminek a neve a paraméterben megadott T típus. A függvény nevére is teszünk megszorítást, ezzel megkapjuk a konstruktorokat, hiszen az osztály nevével megegyező függvények csak konstruktorok lehetnek. Az utolsó attribútummal tovább specializáljuk a függvényt: az argumentumlistának egy elemet adunk meg, egy konstans referenciájú T típust.

A másik része a predikátumoknak a natív, attribútumokkal nem kifejezhető szűrési feltételek megadása. A nyelv bővítményének is felfoghatók, amivel nyelvspecifikusan lehet további megszorításokat elérhetővé tenni a felhasználó számára. Ilyenek például a C++ nyelvben függvényekhez használt *const* vagy *virtual* kulcsszó. Ezek a predikátumokat értelem szerűen csak függvények lekérdezésében alkalmazhatók.

Ezen felül olyan predikátumok is ide tartoznak, amelyek nem szigorúan az entitásokhoz köthető tulajdonságok. Nagyon jó példa erre a változóknál található. Egy változó hivatkozásánál általában sokkal fontosabb, hogy mikor íródik az adott változó, és az olvasás csak mellékes. Mindkét esetre (írás, olvasás) a nyelvnek beépített predikátuma létezik.

4. fejezet

Konklúzió

Irodalomjegyzék

- [1] T. Cohen, J. Y. Gil, and I. Maman. JTL: The Java tools language. In OOPSLA, 2006.
- [2] O. de Moor, M. Verbaere, and E. Hajiyeu. Keynote address: .QL for source code analysis. In SCAM, 2007.
- [3] C. De Roover, C. Noguera, A. Kellens, and V. Jonckers. The SOUL tool suite for querying programs in symbiosis with Eclipse. In PPPJ, 2011.
- [4] D. Janzen and K. De Volder. Navigating and querying code without getting lost. In AOSD, 2003.
- [5] M. Martin, V. B. Livshits, and M. S. Lam, “Finding application errors and security flaws using PQL: a program query language,” in Object Oriented Programming, Systems, Languages and Applications (OOPSLA’07). ACM, 2005, pp. 365–383.
- [6] Bray, Tim. ”The javascript object notation (json) data interchange format.” (2014).