

DOPYTOVANIE DO VEĽKÝCH WEBOVÝCH ÚLOŽÍSK

MATEJ MARCOŇÁK

VEDÚCI PRÁCE: ING. KAROL RÁSTOČNÝ

27. 6. 2013

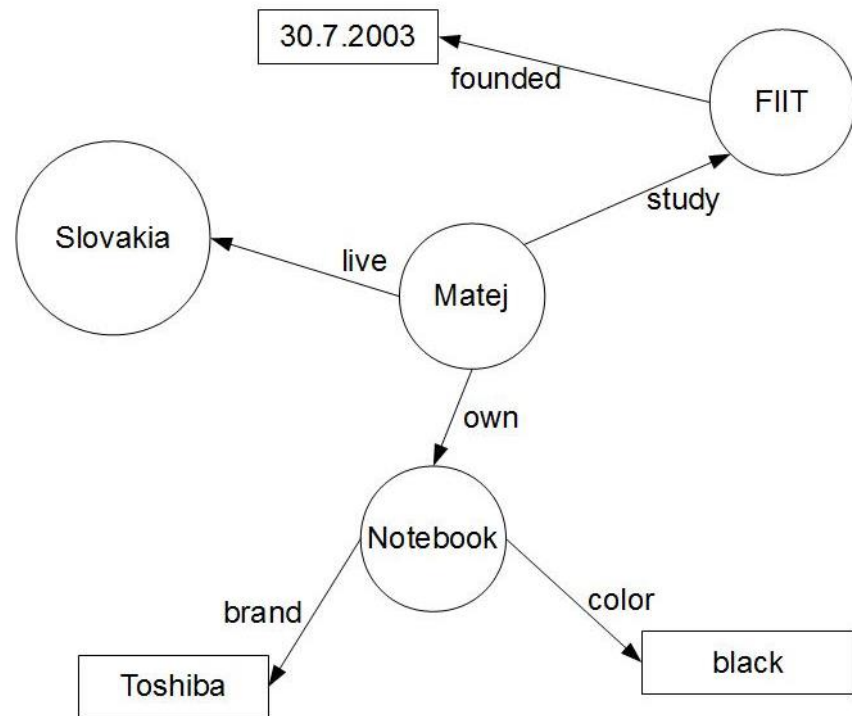
A solid blue horizontal bar spanning the entire width of the slide at the bottom.

Ciele projektu

- ❖ Analyzovať možnosti distribuovaného vyhodnocovania pokročilejších funkcií jazyka SPARQL
- ❖ Navrhnuť a overiť navrhnutú metódu na distribuované vyhodnocovanie netriviálnych SPARQL dopytov pomocou pracovného rámca MapReduce

RDF

- ❖ dátový model
- ❖ orientovaný graf
- ❖ RDF trojice



SPARQL

❖ Dopytovací jazyk

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX oa: <http://www.w3.org/ns/openannotation/core/>
PREFIX typ: <http://purl.org/dc/terms>
SELECT ?tag ?version ?author
WHERE {
    ?tag rdf:type oa:Annotation.
    ?tag oa:annotator ?author.
    ?tag typ:hasVersion ?version.
    FILTER(?version >10)
}
ORDER BY ?version
```

Existujúce riešenia

- ❖ RDF úložiská
- ❖ SQL orientované
- ❖ MapReduce orientované

Návrh riešenia

- ❖ Založený na metóde iteratívneho MapReduce
 - ❖ zníženie počtu potrebných JOIN operácií
- ❖ Zameranie sa na operátory SPARQL
 - ❖ FILTER

```
PREFIX typ: <http://purl.org/dc/terms>
SELECT ?tag ?version
WHERE {
    ?tag typ:hasVersion ?version.
    FILTER(?version >10
           && ?version <20)
}
```

Rozdelenie vyhodnocovania operátora

❖ metóda Map

- ❖ údaje priamo z databázy
- ❖ obmedzené vyhodnocovanie

❖ metóda Reduce

- ❖ príprava údajov na spájanie

❖ metóda Finalize

- ❖ Spájanie RDF trojíc

Identifikované problémy

- ❖ Vzťahy medzi jednotlivými členmi výrazu
- ❖ Počet rôznych premenných v člene výrazu

Vyhodnocovanie v metóde Map

FILTER(?a >10 && ?b <20)

(?a >10)



(?b <20)



FILTER(?a>10 && (?b <20 ||
?c>5))

(?a >10)



(?b <20 || ?c>5)



Vyhodnocovanie v metóde Finalize

```
FILTER(?a >10 && ?b <20  
      && ?c >5)
```

↓

```
1.a,b (?a >10 && ?b <20)
```

↓

```
2.a,b,c (?a >10 && ?b <20  
&& ?c >5)
```

```
FILTER(?a >10 || ?c >5)
```

↓

```
(?a >10 || ?c >5)
```

Optimalizácia výrazu FILTER

- ❖ prepis výrazu do CNF
 - ❖ Distributívne zákony

Experiment

- ❖ Na základe nárastu údajov v databáze
- ❖ Porovnávanie rôznych metód
 - ❖ v metóde Finalize
 - ❖ V spojení metód Map a Finalize
 - ❖ V spojení metód Map a Finalize s optimalizáciou CNF

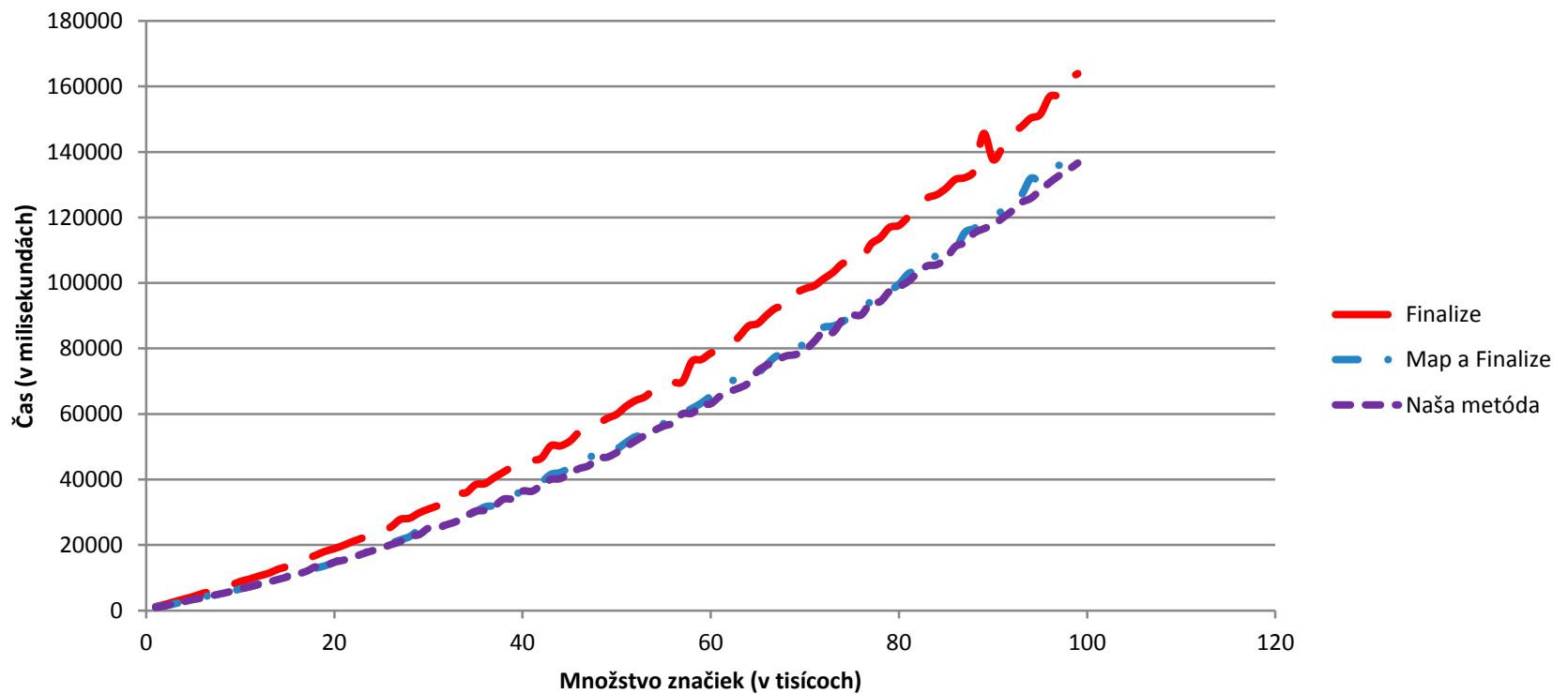
Testované výrazy FILTER

```
FILTER(?author >= <http://concreteauthor.com/author>
    && ?author != <http://concreteauthor.com/Matej>)
}
```

```
FILTER((?author >= <http://concreteauthor.com/author>
    && ?tag != <http://concreteannotation.com>)
|| ?author != <http://concreteauthor.com/Matej> )
}
```

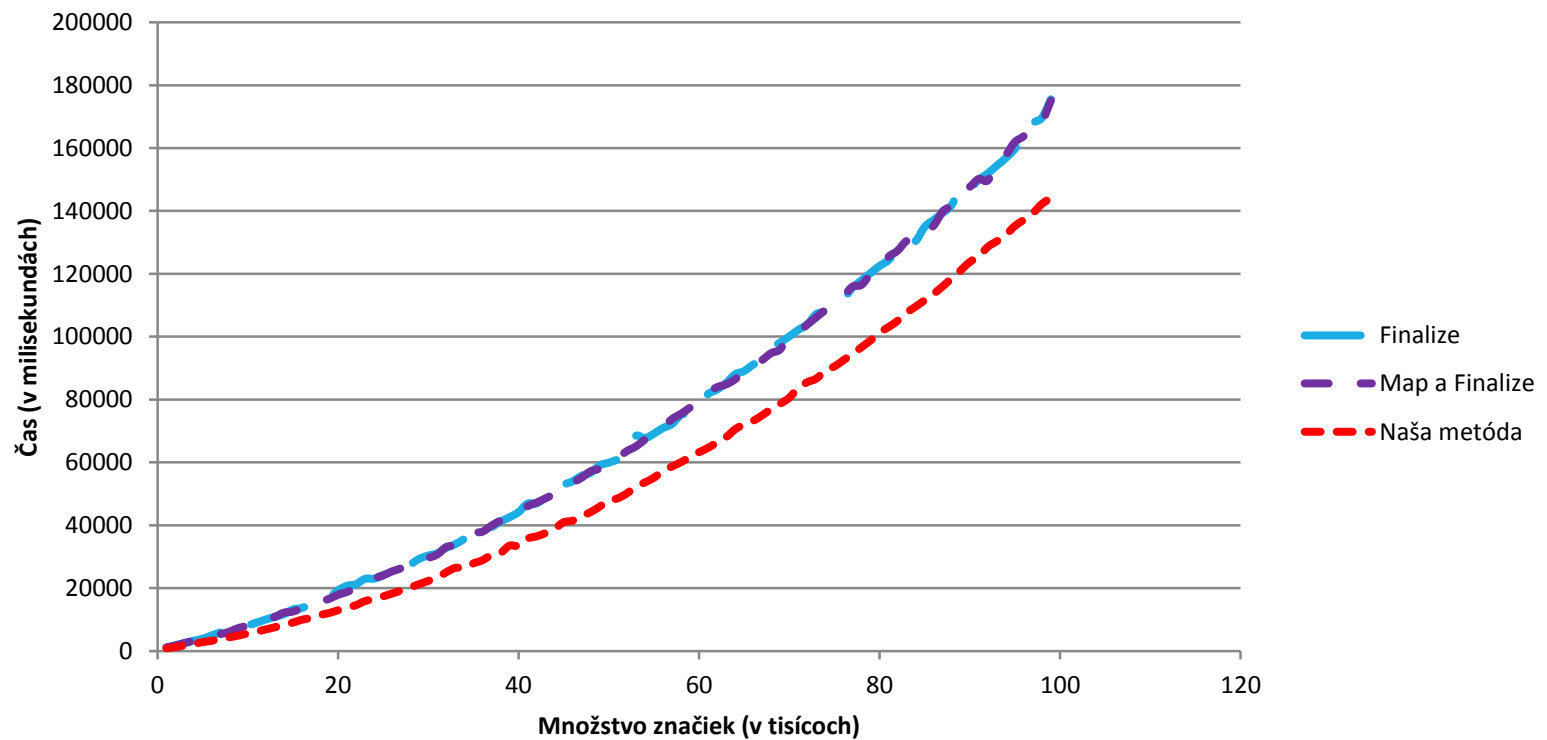
Testovanie 1

Porovnanie metód vzhľadom na 1. dopyt



Testovanie 2

Porovnanie metód vzhľadom na 2. dopyt



Zhodnotenie

- ❖ MapReduce orientované
- ❖ Rozdelenie vyhodnocovania
 - ❖ metóda Map
 - ❖ metóda Finalize
- ❖ Optimalizácia
 - ❖ CNF