

Vyhľadávanie v zdrojových kódach so zohľadnením kontextu

Pavol Zbell

vedúca: prof. Ing. Mária Bieliková, PhD.

27. 6. 2013

konzultant: Ing. Eduard Kuric

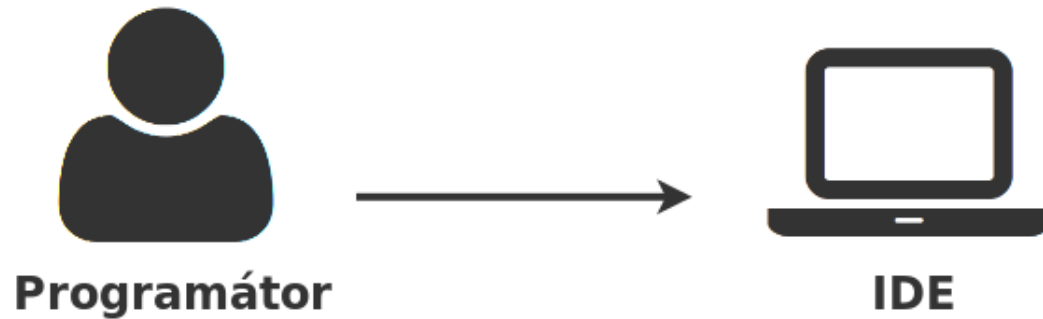
Motivácia

- **Problém:** nedostatočná redukcia priestoru výsledkov vyhľadávania
- **Cieľ:** zjednodušiť, spresniť a sprehľadniť vyhľadávanie
- **Predpoklad:** zmeny zdrojového kódu v určitom čase spolu súvisia
- **Riešenie:** *lokalizácia konceptov* (angl. features) na základe aktivity programátora
- Granularita – *fragment* zdrojového kódu (metóda)

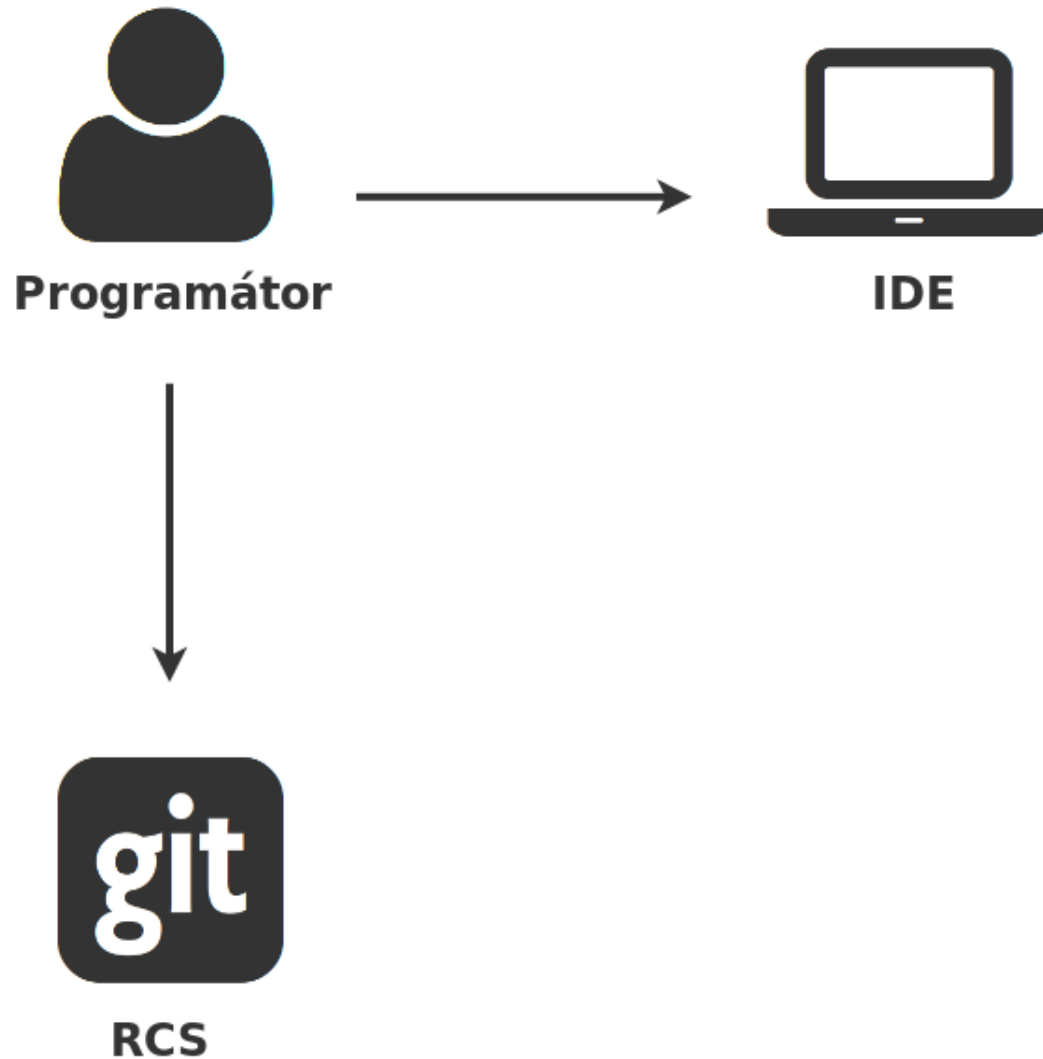
Prostredie

- Významne ovplyvnilo riešenie
- PerConIK, AST-RCS
- Eclipse, Java, git
- ElasticSearch
- Eclipse JDT, Eclipse DOM
- INTT

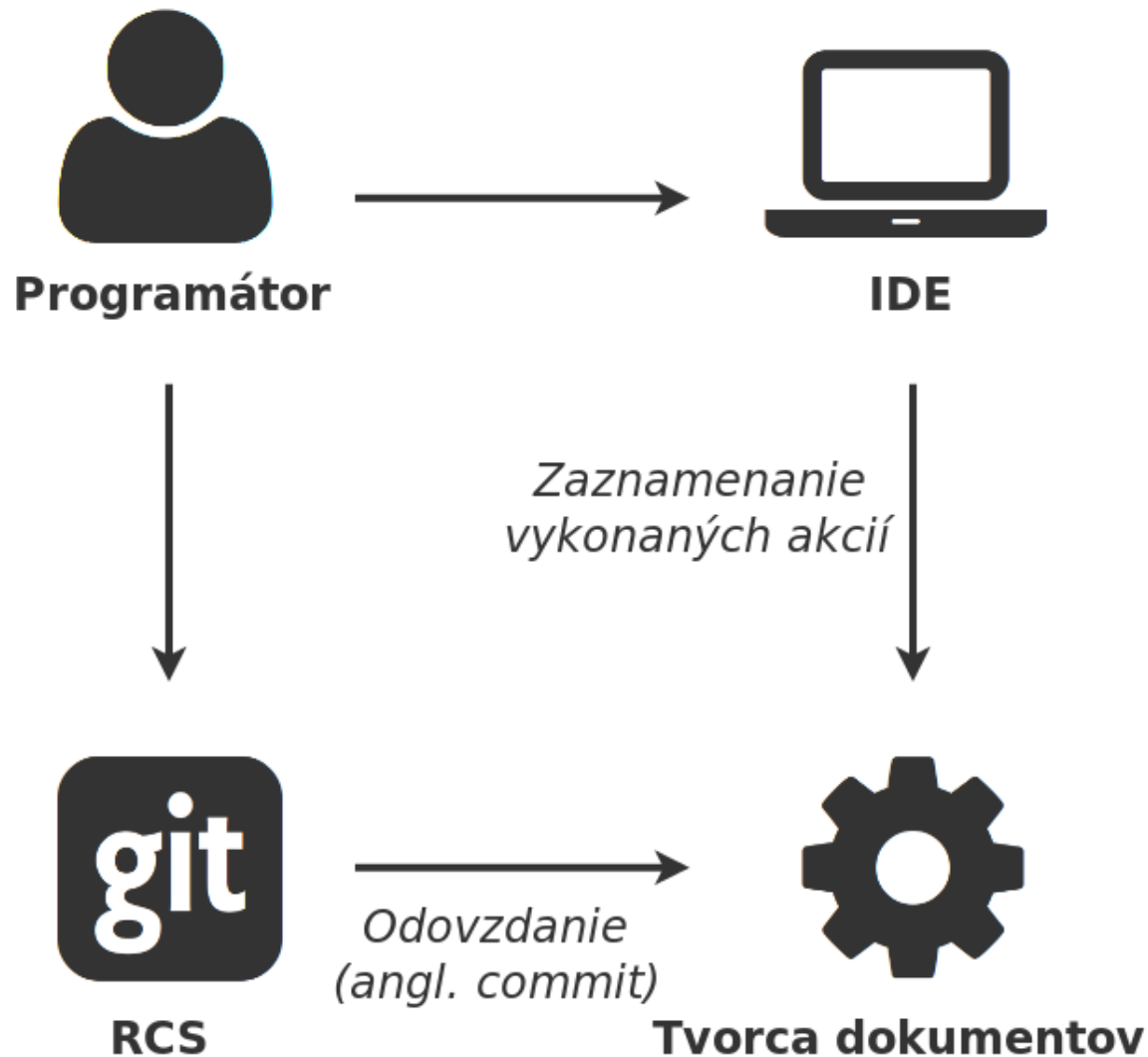
Prehľad metódy



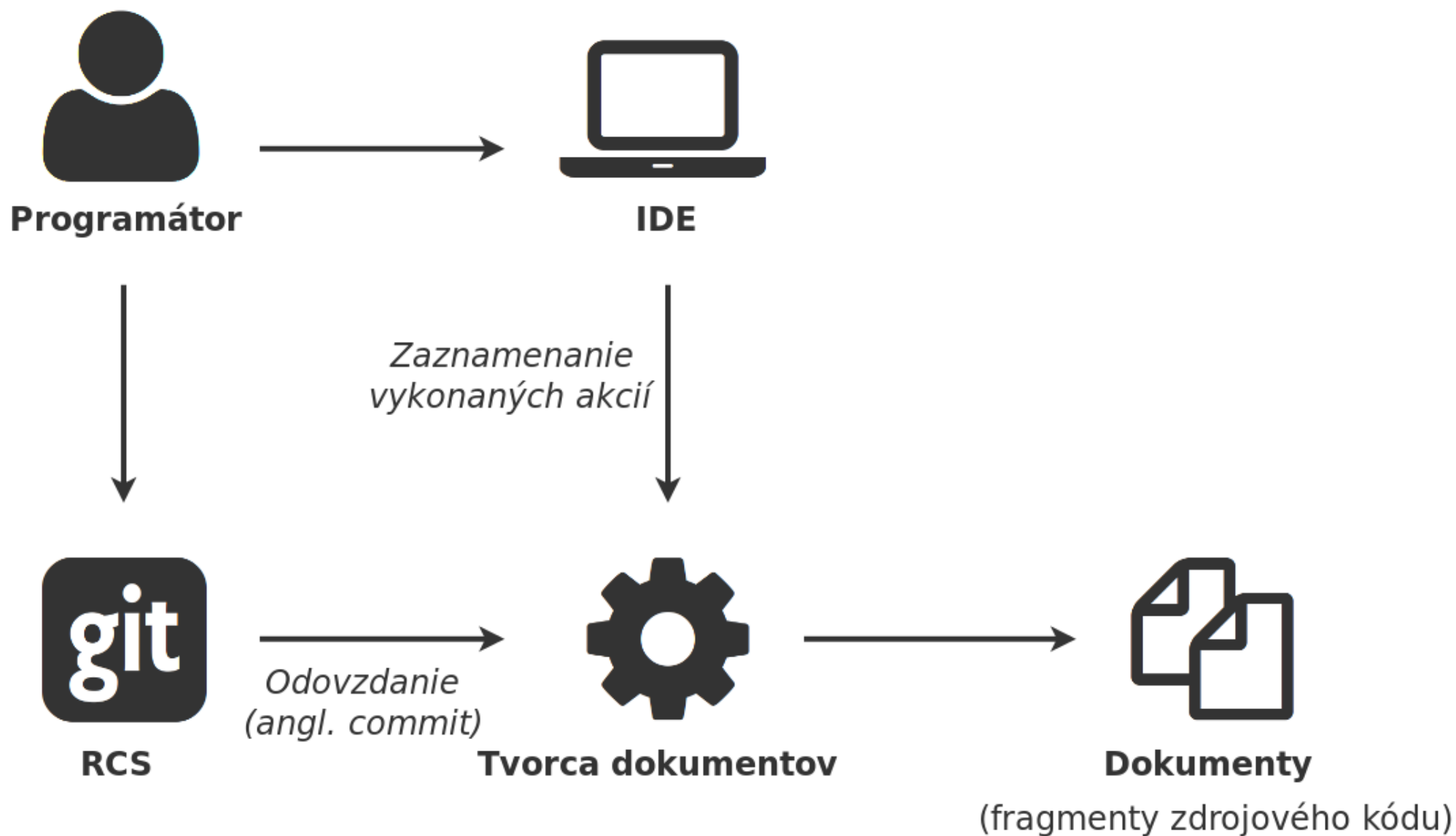
Prehľad metódy



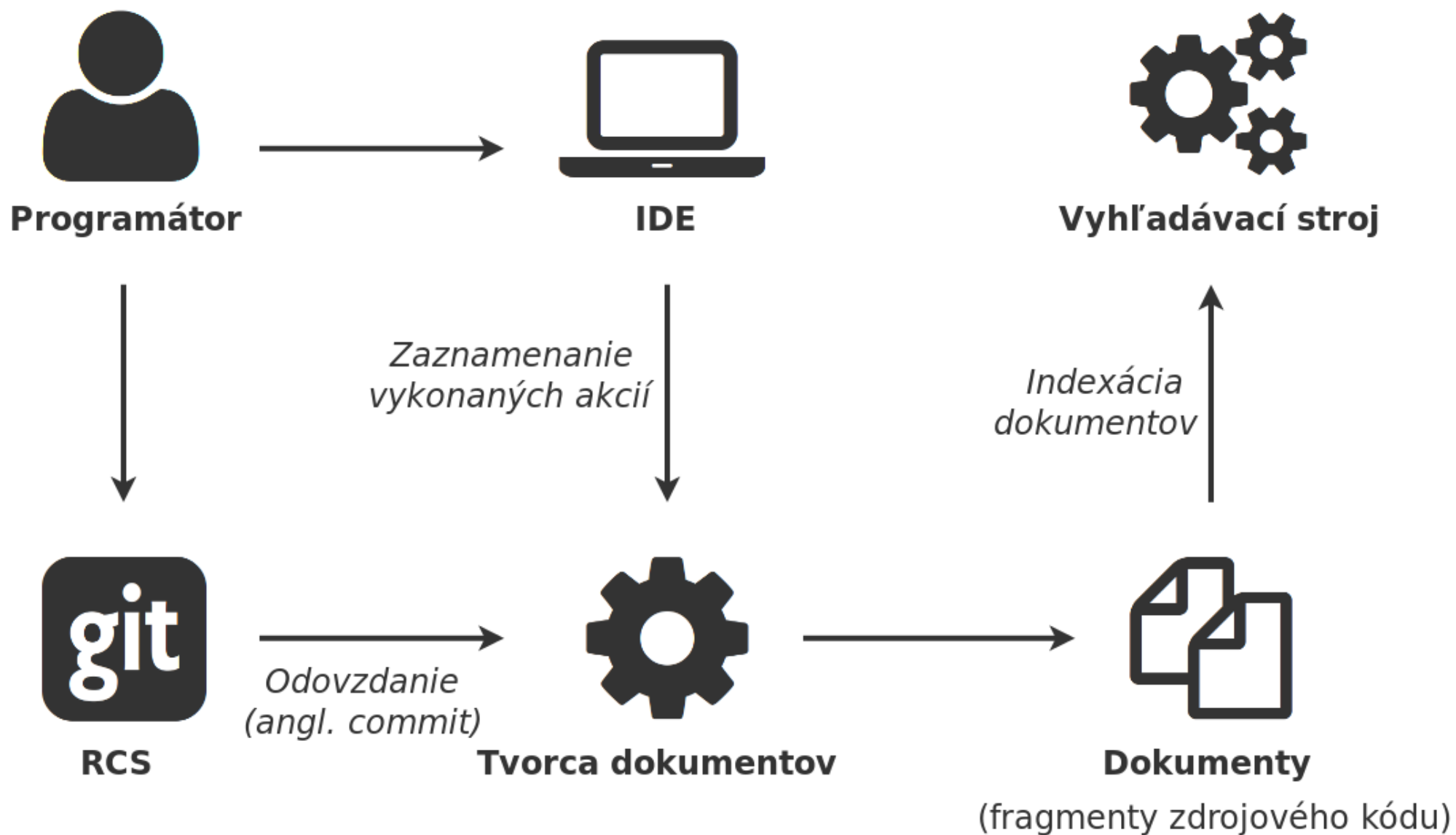
Prehľad metódy



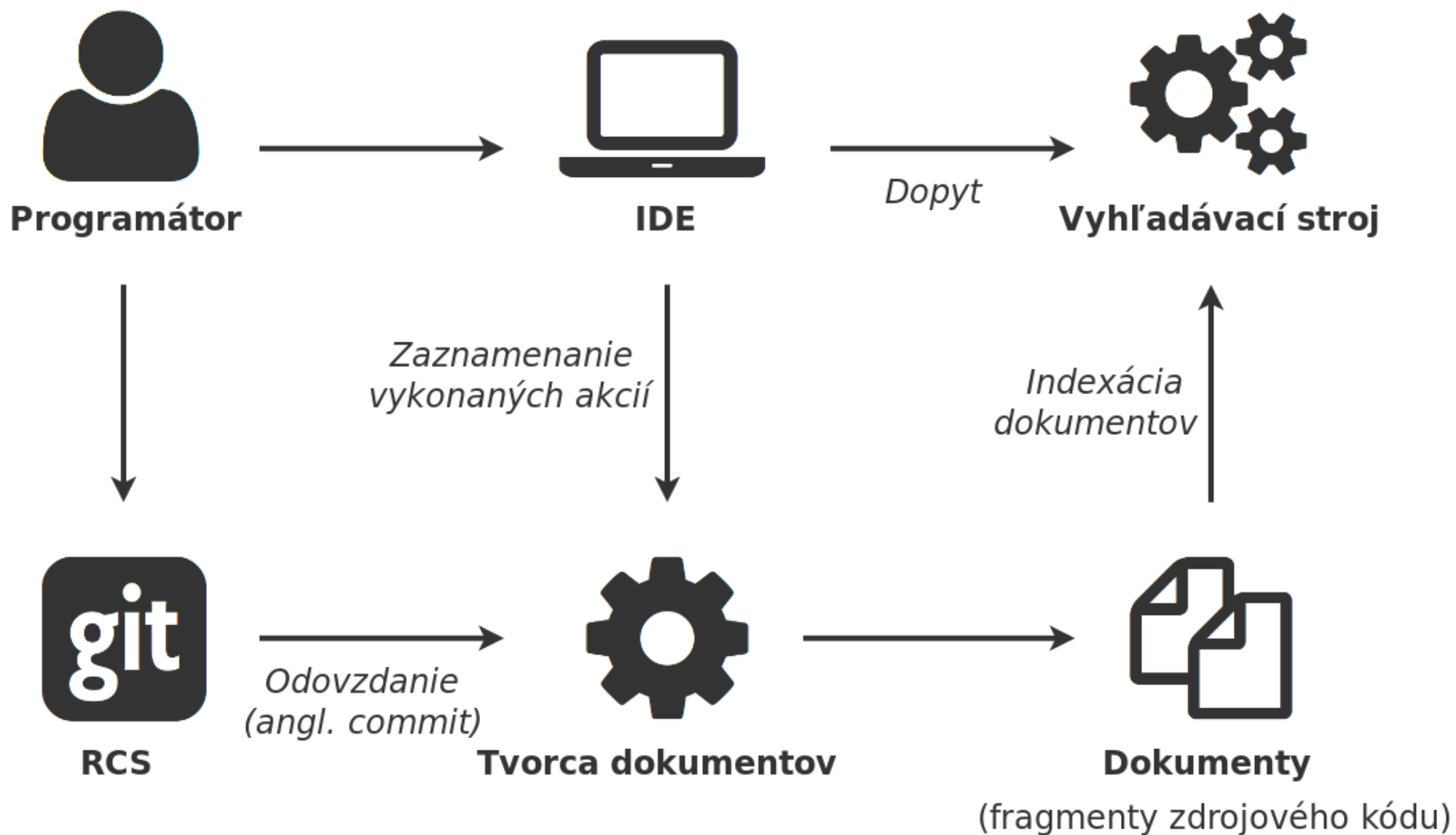
Prehľad metódy



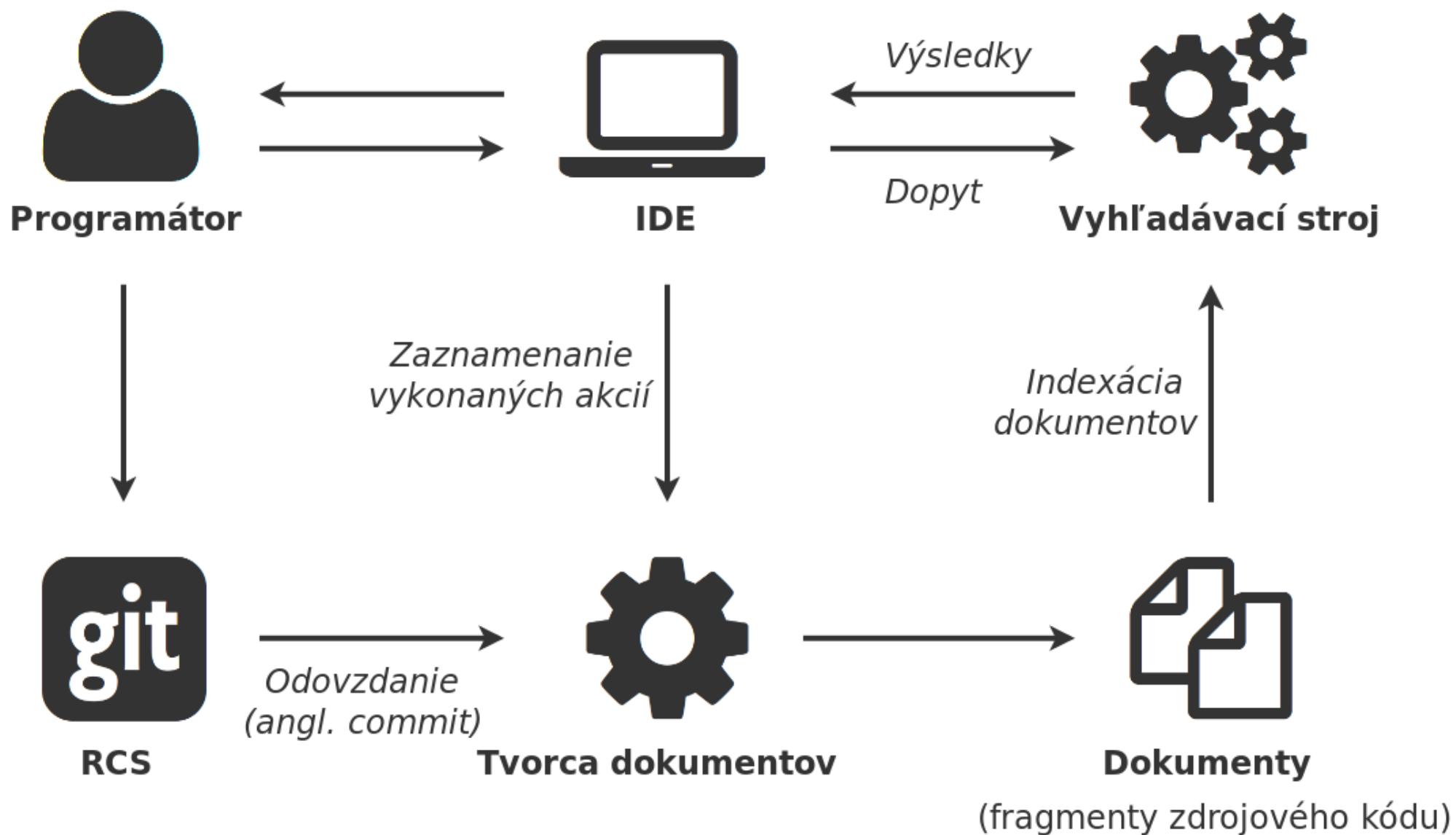
Prehľad metódy



Prehľad metódy



Prehľad metódy

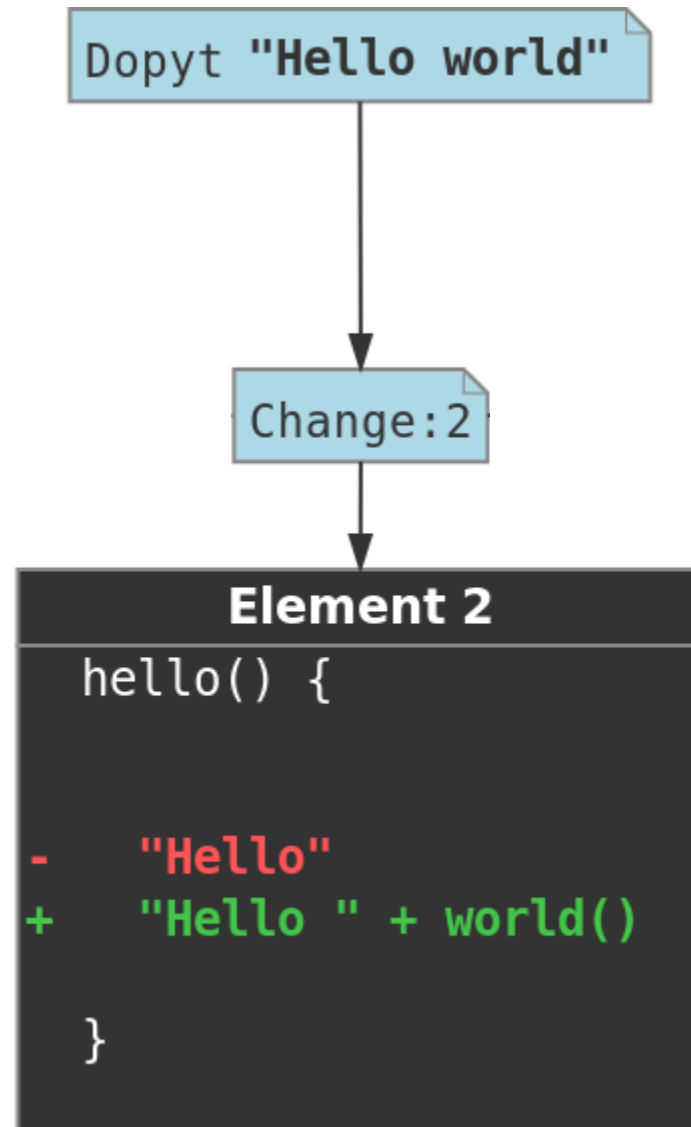


Vyhľadávanie

Dopyt "Hello world"



Vyhľadávanie



Vyhľadávanie

Dopyt "Hello world"

Commit

Change:1

Change:2

Change:3

Element 2

```
hello() {
```

```
-   "Hello"  
+   "Hello " + world()  
  
}
```

Vyhľadávanie

Dopyt "Hello world"

Commit

Change:1

Change:2

Change:3

Element 1

```
+ world() {  
+   "world!"  
+ }
```

Element 2

```
hello() {  
  
-   "Hello"  
+   "Hello " + world()  
  
}
```

Element 3

```
main() {  
  
+   hello();  
  
}
```

Technologické výzvy

- Indexácia a vyhľadávanie
 - mapovanie dokumentov
 - vlastné analyzátory zdrojového kódu
 - vhodná reprezentácia výsledkov
- Zistenie rozdielu dvoch AST
 - transformácia, diff, porovnanie fragmentov
- Jemná granularita sledovania udalostí v IDE
 - sledovanie malých zmien, optimalizácia algoritmov
 - časté syntakticky nesprávne AST

Overenie

- Vyhľadávanie
 - relevancia nájdeneého fragmentu voči dopytu
- Lokalizácia konceptov
 - prínos zohľadnenia zmien zdrojového kódu v čase
- Detailnejší návrh oboch experimentov
 - ciele, priebeh, vyhodnotenie

Changes made right after found match

Main.java on 24. 06. 2013 at 02:28:34

Main.java on 24. 06. 2013 at 02:28:28

Match found in World.java on 24. 06. 2013 at 02:28:15

URI: file:/base/workspace/runtime-EclipseApplication/simple-example/src/World.java

Name Path: _/World/toString

Type Path: CompilationUnit/TypeDeclaration/MethodDeclaration

```
@Override public final String toString(){  
    return "world";  
}
```

Changes made right before found match

World.java on 24. 06. 2013 at 02:28:13

Changes made right after found match

Main.java on 24. 06. 2013 at 02:28:35

URI: file:/base/workspace/runtime-EclipseApplication/simple-example/src/Main.java

Name Path: _/Main/main

Type Path: CompilationUnit/TypeDeclaration/MethodDeclaration

```
public static final void main(final String[] args){  
    System.out.println("Hello " + new World().toString() + "!");  
}
```

Main.java on 24. 06. 2013 at 02:28:28

Match found in World.java on 24. 06. 2013 at 02:28:15

URI: file:/base/workspace/runtime-EclipseApplication/simple-example/src/World.java

Name Path: _/World/toString

Type Path: CompilationUnit/TypeDeclaration/MethodDeclaration

```
@Override public final String toString(){  
    return "world";  
}
```

Prínos

- Vyhľadanie najrelevantnejších metód
- Preskúmavanie zdrojového kódu
- Pohľad na evolúciu zdrojového kódu (ako celku)
- Obohatenie Eclipse DOM API na prácu s AST
- Možné rozšírenia:
 - pohľad na evolúciu konkrétneho fragmentu
 - pridanie metadát, užšia integrácia so službou AST-RCS

Zhrnutie

- Zdrojový kód modifikovaný v čase blízko seba spolu súvisí
- Zaznamenanie zmien AST v čase a indexácia fragmentov zdrojového kódu
- Lokalizácia konceptov
 - na základe podobnosti dopytu a zdrojového kódu
 - vyhľadávaním najrelevantnejších fragmentov
 - prezentovaná ako sled zmien zdrojového kódu
- Pohľad na evolúciu zdrojového kódu

Návrh experimentov

- Vyhľadávanie fragmentov zdrojového kódu
 - relevancia nájdeného fragmentu voči dopytu
 - experti – programátori, schopní ohodnotiť relevanciu
 - úlohy (modifikácia konceptov v známych projektoch)
 - rozdelenie programátorov na dve skupiny
 - porovnanie priemerného počtu dopytov
 - kvalitatívna analýza dopytov
 - porovnanie oboch skupín (ako výsledky ovplyvňujú ďalšie dopyty)
 - experimentovanie s nastaveniami Elasticsearch

Návrh experimentov

- Lokalizácia konceptov
 - prínos zohľadnenia zmien zdrojového kódu v čase
 - dve skupiny programátorov
 - neznámy projekt vytvorený pomocou IDE rozšírenia
 - úlohy (lokalizovanie miesta potrebnej zmeny)
 - kvalitatívne vyhodnotenie pripravených otázok
 - porovnanie oboch skupín (čas riešenia, počet dopytov, ...)
 - experimentovanie s prezentáciou výsledkov