

# Identifikace uzlu ve výsledku dotazu do MDFS

## Úvod

MDFS jako úložiště dat poskytuje vlastní systém pro vyhledávání v uložených datech. Při práci s daty se většinou postupuje stylem „najdi něco a zpracuj to“. První fáze je tedy vyhledávání ve kterém lze rozlišit tři základní přístupy:

- lineární přístup ke všem uzlům
- přístup ke kořenovým uzlům
- dotaz

První způsob poskytuje nestrukturovaný přístup ke všem uzlům což není vždy praktické a proto existuje druhý přístup, který nám řekne, které uzly jsou kořenové (nejsou v roli potomka). Druhým způsobem však není možné přistupovat k zacykleným strukturám, takže taky není obecně použitelný.

Třetí přístup, dotazovací, je nejuniverzálnější, ale taky implementačně nejnáročnější. Při správném dotazu však může nahradit také funkci prvních dvou metod. Pomocí této metody získáme množinu uzlů, které vyhovují vyhledávacímu dotazu.

Typ výsledku je složen ze dvou komponent - rozměru a základního typu. Možnosti v pozici rozměru jsou následující:

- skalární výsledek
- vektor výsledků

Jako základní typ je pak jedna z následujících možností:

- číselná hodnota
- řetězcová hodnota
- uzel

V prvních dvou případech jde o jednoduchou hodnotu která je určena na přímé zpracování. V případě výsledku jako uzlu se však předpokládá, že se nalezený uzel použije jako výchozí bod pro zpracovávající algoritmus. V tomto způsobu použití pak potřebujeme jednoznačně identifikovat uzel pro následující dotazy a metoda identifikace je předmětem tohoto dokumentu.

## Problémy

Jedním z problémů který zabraňuje použití triviální identifikace pořadovým číslem uzlu je abstraktní pojetí služby MDFS. Implementačně pak může vypadat MDFS jako:

- nadstavba nad SQL databází
- nadstavba nad množinou souborů

První implementace může použít triviální identifikaci ale druhá ne, protože by bylo nutno vymyslet mapovací funkci která převede číslo na název souboru a počet uzlů by byl omezen.

## Možnosti

Ke službě MDFS se bude přistupovat přes webové služby (SOAP protokol, XML formát dat), takže se povoluje použití libovolně strukturovaných identifikátorů.

## Řešení

Řešení identifikace spočívá právě v použití volitelně strukturovaných identifikátorů uzlů. Pro názornost definuji identifikátory příkladem - nejprve triviální:

```
<id>123567890</id>
```

a pak strukturovaný:

```
<id>
  <file>/dir/filename.xml</file>
  <query>//nodename</query>
  <node>123</node>
</id>
```