

Rozhraní pro fulltextové vyhledávání v CUL



Operace vyhledej a vyhledejDalsi jsou dostupné od verze CUL 1.2.

1. Rozhraní pro fulltextové vyhledávání v CUL

Fulltextové vyhledávání v CUL je realizováno pomocí operací "vyhledej" a "vyhledejDalsi". Tyto operace jsou součástí webových služeb DigitalniDokumentService na CULWS rozhraní.

Scénář případu užití

0. Případ užití začíná v okamžiku, kdy primární systém odešle pomocí operace "vyhledej" hledaný řetězec znaků.

1. CUL provede hledání ve všech indexovaných metadatech (název, popis,...) a ve fulltextu (binární data digitálního dokumentu).

1.1 KDYŽ je v requestu operace "vyhledej" zadán parametr "primarniSystem", omez množinu prohledávaných digitálních dokumentů na digitální dokumenty pouze od daného primárního systému.

2. CUL vyhodnotí práva uživatele Alfresco na zobrazení nalezených digitálních dokumentů.

2.1 KDYŽ primární systém přistupující do CUL pomocí uživatele Alfresco nemá práva na všechny nalezené digitální dokumenty, omez množinu získanou v bodě 1 dle uživatelských práv.

3. CUL vrátí primárnímu systému výslednou množinu GUID digitálních dokumentů seřazených dle relevance.

4. Případ užití končí.

1.1. Operace "vyhledej"

Operace "vyhledej" slouží pro hledání digitálních dokumentů dle zadaného řetězce a primárního systému (volitelně). Ve hledaném řetězci lze libovolný počet znaků nahradit symbolem * (uplatnění zejména v názvech souborů). Hledání probíhá ve všech indexovaných metadatech (název, popis,...) a ve fulltextu (binární data digitálního dokumentu).

Pro fulltextové vyhledávání jsou podporovány Office dokumenty, xml, pdf, txt,... Lze nastavit počet vrácených výsledků (defaultně 100). Maximální celkový počet výsledků je omezen nastavením CUL, které se konfiguruje pomocí webového rozhraní Alfresco (defaultně 1000). Vrácené výsledky jsou řazeny dle relevance. Dokumentace a příklady Request/Response jsou umístěny na rozcestník rozhraní CULWS.

Pozn 1.: Pokud by bylo třeba hledat pouze ve fulltextu, lze jednoduše upravit. Nyní se hledá ve fieldu ALL viz <http://wiki.alfresco.com/wiki/Search>.

Pozn 2.: Operace "vyhledej" hledá pouze celá slova (opomineme-li speciální zástupné znaky - [wildcard](#)), která jsou dopředu zaindexována (Lucene indexy), aby bylo možné je co nejrychleji vyhledat. Česká diakritika nepředstavuje žádný problém.

Pozn 3.: Primární systémy Verso a iFIS mají přidělena práva Contributor na prostor CUL. To umožňuje prohlížet vše, ale měnit (update, remove) pouze své dokumenty.

1.1.1. Omezení množiny prohledávaných digitálních dokumentů

Množinu prohledávaných digitálních dokumentů lze omezit dvěma způsoby.

a) Parametr "primarniSystem" v requestu operace "vyhledej"

Pomocí parametru "primarniSystem" je možné zadat identifikátor primárního systému, pro hledání dokumentů pouze od konkrétního primárního systému. Pokud není parametr zadán, hledá se nad všemi digitálními dokumenty. Operace "vyhledej" se chová stejným způsobem jako vyhledávání přímo přes webového klienta Alfresco.

b) Omezení pomocí práv uživatele v Alfresco

Každý primární systém má v Alfresco přiděleného uživatele. Pomocí nastavení práv uživatele (provádí se přes webové rozhraní Alfresco) lze definovat, která data může daný primární systém prohlížet, vyhledávat, popřípadě upravovat.

1.2. Operace "vyhledejDalsi"

Pro načtení dalších výsledků (další stránky) slouží operace "vyhledejDalsi". Jako vstup využívá identifikátor hledání (dotazToken) vrácený předchozím hledáním. Pokud v jednu chvíli přistoupí k operaci "vyhledej" více primárních systémů, může existovat více platných dotazToken. Operace "vyhledejDalsi" není tedy omezena z hlediska počtu přístupujících primárních systémů v jednom okamžiku.

Platnost dotazToken je však časově omezena (defaultně 5 minut). Časovou platnost lze nakonfigurovat přímo v Alfresco (konfigurační soubor ehcache-default.xml):

```
<cache
name="org.alfresco.repo.webservices.querySessionSharedCache"
maxElementsInMemory="1000"
eternal="false"
timeToLiveSeconds="300"
overflowToDisk="false"
statistics="false"
/>
```

Dokumentace a příklady Request/Response jsou umístěny na rozcestník rozhraní CULWS.