

Brug af facetter fra Opensearch

Dette dokument er et supplement og uddybelse til den tekniske beskrivelse af Opensearch på opensource.dbc.dk.

Eksemplerne i dokumentet kan afprøves i eksempelklienten for Opensearch via opensearch.addi.dk men vil ikke altid give præcis samme resultat som eksemplerne i dette dokument, da poster i brønden løbende ændres.

Facet-parametre i Opensearch.....	1
Angivelse af facetnavne.....	2
Hvor mange termer pr. facetnavn.....	3
Sortering af termer.....	3
Minimum hits.....	4
Udlever facetter med eller uden poster.....	4
Trinvis indsnævring af søgeresultat via facetter.....	6
Udvælgelse af bestemte facet-termer.....	8

Facet-parametre i Opensearch

Når man sender en søgning igennem Opensearch har man mulighed for at få udleveret facetter (i Opensearch 'facets').

Man angiver **et eller flere facetnavne** (facetName) som man vil have udleveret, f.eks. facet.creator (ophav) og har derudover mulighed for at specificere:

- Hvor mange termer pr. facetnavn man vil have udleveret ('numberOfTerms')
- Hvordan man vil have facetterne sorteret ('facetSort') – man kan sortere alfabetisk eller efter antal forekomster
- Hvor mange hits der minimum skal være i hver enkelt facet for at den bliver udleveret af servicen ('facetMinCount')

Man kan bede om facetter **sammen med poster** eller **for sig selv** alt efter hvad man gerne vil med sin grænseflade.

Angivelse af facetnavne

Med `facetName` angiver man hvilke typer af facetter man gerne vil have udleveret. Hvilke facetnavne der kan bruges i Opensearch kan ses på <http://www.danbib.dk/brondindekser> (man skal bruge dem der starter med 'facet' efterfulgt af et punktum og en specificering af facetten).

I dette eksempel ønskes udleveret facetter inden for materialetypekategori (`facet.typeCategory`):

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:numberOfTerms>10</open:numberOfTerms>
        <open:facetSort>count</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
        <open:facetName>facet.typeCategory</open:facetName>
      </open:facets>
      <open:start>1</open:start>
      <open:stepValue>10</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
```

Man kan bede om flere facetnavne på én gang ved blot at gentage parameteren `facetName`, f.eks. som her hvor der ønskes facetter inden for både `facet.typeCategory` og `facet.type`:

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:numberOfTerms>10</open:numberOfTerms>
        <open:facetSort>count</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
        <open:facetName>facet.typeCategory</open:facetName>
        <open:facetName>facet.type</open:facetName>
      </open:facets>
      <open:start>1</open:start>
      <open:stepValue>10</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
```

De øvrige parametre der er specificeret for udlevering af facetter (`numberOfTerms`, `facetSort` og `facetMinCount`) gælder for alle facetnavne der er bedt om. Det vil sige, at man ikke kan bede om f.eks. én sortering for `facet.typeCategory` og en anden for `facet.type` med mindre man beder om de to facetnavne i to separate kald til Opensearch.

Hvor mange termer pr. facethavn

Hvor mange termer man ønsker udleveret pr. facethavn kan man angive med 'numberOfTerms' under elementet facets.

Denne søgning vil f.eks. udlevere **fem** termer inden for facethavn 'facet.typeCategory':

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:numberOfTerms>5</open:numberOfTerms>
        <open:facetSort>count</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
        <open:facetName>facet.typeCategory</open:facetName>
      </open:facets>
      <open:start>1</open:start>
      <open:stepValue>10</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Hvor mange termer der skal udleveres pr. facethavn er et spørgsmål om, hvor mange termer man vil præsentere brugeren for i grænsefladen. Man kan f.eks. vælge at præsentere fem termer pr. facethavn til en begyndelse, og give brugeren mulighed for at vælge at udvide dette til ti eller flere, i hvilket tilfælde der sendes en ny forespørgsel til Opensearch hvor numberOfTerms er ændret fra '5' til '10'.

Sortering af termer

Man kan selv vælge i hvilken rækkefølge man ønsker termene inden for de enkelte facethavne sorteret. Der er to muligheder; enten alfanumerisk (facetSort=index) eller efter antal forekomster (facetSort=count). Hvis man ikke angiver en facetSort vil der automatisk sorteres efter antal forekomster.

I dette eksempel er valgt at de fem termer der skal udleveres inden for facethavnet facet.typeCategory, skal være sorteret med den term der forekommer oftest i søgeresultatet først. Det vil sige at man vil få udleveret de fem hyppigst forekommende termer for typeCategory inden for de poster der hittes med den angivne query:

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:numberOfTerms>5</open:numberOfTerms>
        <open:facetSort>count</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
      </open:facets>
    </open:searchRequest>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

```
<open:facetName>facet.typeCategory</open:facetName>  
</open:facets>  
<open:start>1</open:start>  
<open:stepValue>10</open:stepValue>  
</open:searchRequest>  
</SOAP-ENV:Body>
```

Havde man i stedet valgt `facetSort=index`, da ville man have fået udleveret fem termer i alfanumerisk rækkefølge uanset hvor mange hits der er under hver enkelt term, dog således at termer der ikke har hits inden for søgeresultatet, ikke udleveres.

Minimum hits

Med `facetMinCount` parameteren kan man vælge hvor mange poster inden for søgeresultatet der minimum skal falde inden for hver term, for at termen udleveres i facet-resultatet. Hvis man ikke angiver noget `facetMinCount` medtages alle termer hvor der er minimum én forekomst inden for søgeresultatet.

Man skal være varsom med at sætte `facetMinCount` til et meget højere tal end 1 da man især ved meget snævre søgninger med få hits risikerer at man slet ikke får udleveret facet-termer.

I dette eksempel er der bedt om facetter med navnet `facet.sheetmusic` (nodetype) og der skal udleveres ti termer sorteret efter forekomst, men kun de termer der har mindst fem forekomster:

```
<?xml version="1.0" encoding="UTF-8"?>  
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"  
xmlns:open="http://oss.dbc.dk/ns/opensearch">  
  <SOAP-ENV:Body>  
    <open:searchRequest>  
      <open:query>klima</open:query>  
      <open:agency>100200</open:agency>  
      <open:profile>test</open:profile>  
      <open:facets>  
        <open:numberOfTerms>10</open:numberOfTerms>  
        <open:facetSort>count</open:facetSort>  
        <open:facetMinCount>5</open:facetMinCount>  
        <open:facetName>facet.sheetMusic</open:facetName>  
      </open:facets>  
      <open:start>1</open:start>  
      <open:stepValue>1</open:stepValue>  
    </open:searchRequest>  
  </SOAP-ENV:Body>  
</SOAP-ENV:Envelope>
```

Ovenstående eksempel kunne f.eks. bruges i tilfælde hvor man ønsker i første omgang kun at vise bestemte facetter (her `nodetype`) i et søgeresultat, hvis der er en hvis mængde af poster der falder inden for den valgte kategori.

Udlever facetter med eller uden poster

I forespørgslen til Opensearch kan man sætte `start=0` og `stepValue=0` samtidig med at man beder om at få udleveret facetter, for at få et resultat der udelukkende indeholder gruppering af søgeresultatet inden for de facetter man har bedt om, og ikke indeholder poster. Det kan f.eks. være som en del af en parallel-

forespørgsel til servicen; én hvor der kun udleveres poster og én hvor der kun udleveres facetter. Hvis man har brug for at præsentere posterne ét sted og facetterne et andet sted på grænsefladen, kan parallelforespørgselsmetoden måske være nyttig.

Denne forespørgsel til Opensearch udleverer kun facetter og ingen poster:

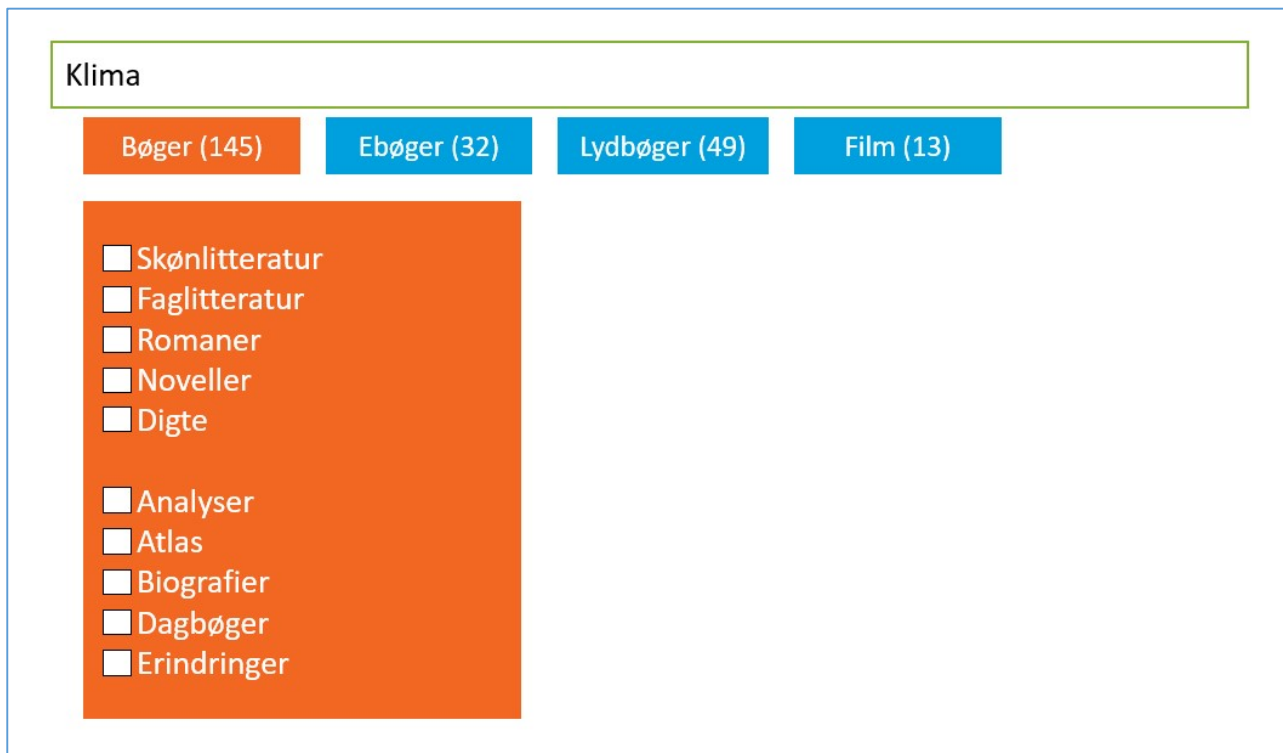
```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:numberOfTerms>10</open:numberOfTerms>
        <open:facetSort>count</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
        <open:facetName>facet.typeCategory</open:facetName>
      </open:facets>
      <open:start>0</open:start>
      <open:stepValue>0</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
```

Mens denne forespørgsel er samme query udført, hvor der ønskes poster retur men ingen facetter:

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:start>1</open:start>
      <open:stepValue>10</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
```

Trinvis indsnævring af søgeresultat via facetter

Eksempel: Grupper søgeresultat efter materialekategori inden visning af poster, så bruger kan klikke sig frem til et overskueligt sæt.



En søgning – her på 'klima' - hvor der ikke udleveres poster (start=0, stepValue=0) men kun facettering af resultatet i første omgang ud fra materialetypekategorier (facetName=facet.typeCategory):

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:numberOfTerms>10</open:numberOfTerms>
        <open:facetSort>count</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
        <open:facetName>facet.typeCategory</open:facetName>
      </open:facets>
      <open:start>0</open:start>
      <open:stepValue>0</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Ovenstående giver et søgeresultat f.eks. som dette:

```
<result>
  <hitCount>22015</hitCount>
  <collectionCount>0</collectionCount>
  <more>>true</more>
  <facetResult>
    <facet>
      <facetName>facet.typeCategory</facetName>
      <facetTerm>
        <frequency>120</frequency>
        <term>bøger</term>
      </facetTerm>
      <facetTerm>
        <frequency>61</frequency>
        <term>ebøger</term>
      </facetTerm>
      <facetTerm>
        <frequency>53</frequency>
        <term>lydbøger</term>
      </facetTerm>
      <facetTerm>
        <frequency>14</frequency>
        <term>film</term>
      </facetTerm>
    </facet>
  </facetResult>
</result>
```

Ud fra valg af materialetypekategori, kan der præsenteres et yderligere indsnævret søgeresultat fordelt efter f.eks. litterær form ('facet.literaryForm') og indholdets form ('facet.form').

Her er som eksempel et kald hvor der er valgt 'bøger' (query er udvidet fra 'klima' til 'klima AND facet.typeCategory=bøger') og der skal udleveres facettering af resultatet ud fra to facetter; 'facet.literaryForm' og 'facet.form'. Denne gang vælges, at facetterne skal sorteres alfanumerisk (facet.sort=index) i stedet for med højeste forekomst først:

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima AND facet.typeCategory="bøger"</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:facets>
        <open:facetSort>index</open:facetSort>
        <open:facetMinCount>1</open:facetMinCount>
        <open:facetName>facet.literaryForm</open:facetName>
        <open:facetName>facet.form</open:facetName>
      </open:facets>
      <open:start>0</open:start>
      <open:stepValue>0</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Ovenstående vil give et søgeresultat der bl.a. indeholder følgende:

```
<facetResult>
  <facet>
    <facetName>facet.literaryForm</facetName>
    <facetTerm>
      <facetTerm>
        <frequency>18</frequency>
        <term>digte</term>
      </facetTerm>
      <frequency>2388</frequency>
      <term>faglitteratur</term>
    </facetTerm>
    <facetTerm>
      <frequency>18</frequency>
      <term>noveller</term>
    </facetTerm>
    <facetTerm>
      <frequency>73</frequency>
      <term>romaner</term>
    </facetTerm>
    <facetTerm>
      <frequency>139</frequency>
      <term>skønlitteratur</term>
    </facetTerm>
  </facet>
  <facet>
    <facetName>facet.form</facetName>
    <facetTerm>
      <frequency>2</frequency>
      <term>analyser</term>
    </facetTerm>
    <facetTerm>
      <frequency>15</frequency>
      <term>atlas</term>
    </facetTerm>
    <facetTerm>
      <frequency>8</frequency>
      <term>biografier</term>
    </facetTerm>
    <facetTerm>
      <frequency>1</frequency>
      <term>dagbøger</term>
    </facetTerm>
    <facetTerm>
      <frequency>14</frequency>
      <term>erindringer</term>
    </facetTerm>
    [afkortet]
  </facet>
```

De nye facetter kan præsenteres i grænsefladen til yderligere indsnævring af bøger om klima, jf. eksempel-illustrationen øverst.

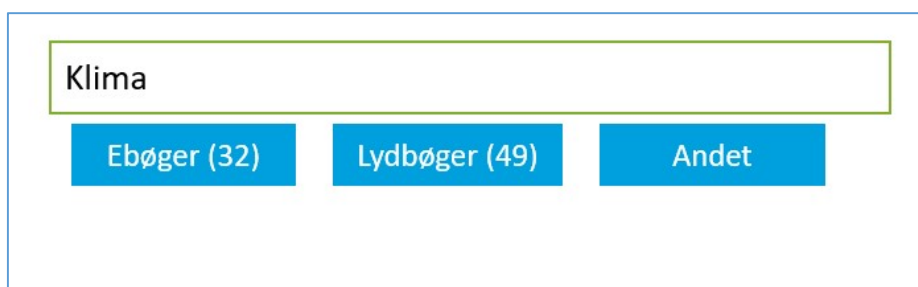
Udvælgelse af bestemte facet-termer

De forskellige facetter der er tilgængelige i Opensearch, kan have enten en længere og nærmest uudtømmelig liste af facet-termer der kan præsenteres i grænsefladen, eller en kortere og afgrænset liste af termer. Det fremgår af dokumentationen af facet-indekser på <http://www.danbib.dk/brondindekser> om

det er den ene eller den anden type, ved at der er beskrevet enten 'mulige værdier' (afgrænset liste) eller 'eksempler' (uudtømmelig liste) på facet-termer.

Der kunne måske være situationer, hvor man i grænsefladen kun vil udnytte enkelte af de mulige værdier, f.eks. at man altid kun vil vise termerne 'ebøger' og 'lydbøger' fra facet.typeCategory og ikke vil vise 'bøger', 'film', 'musik' osv. I sådanne tilfælde skal udvælgelsen ske af grænsefladen selv på baggrund af det returnerede resultat fra Opensearch. Det samme gælder, hvis man gerne vil præsentere facet-termerne i en anden rækkefølge end enten alfa-numerisk eller efter antal forekomster i søgeresultatet.

Hvis man udvælger bestemte facet-termer til visning i grænsefladen og dermed udelader andre, kan man f.eks. vælge at konstruere en 'Andet'-kategori, som indeholder alt det man har sorteret fra og de poster der ikke har værdier inden for den valgte facet. Blot kan man ikke fortælle brugeren, hvor stor en del af søgeresultatet der ligger inden for denne kategori. Eksemplet kan illustreres således:



Hvis brugeren vælger dette 'Andet', skulle der sendes en forespørgsel til Opensearch indeholdende den oprindelige query kombineret med at søgeresultatet nu IKKE må indeholde de facet-termer man har udvalgt inden for den facet man har valgt sine facet-termer fra.

Her er som eksempel et kald til Opensearch der viser hvilken forespørgsel der skulle genereres, hvis brugeren valgte kategorien 'Andet' på ovenstående illustration, hvor facet-termerne 'ebøger' og 'lydbøger' er specifikt udvalgt som eneste termer der vises. Query'en 'klima' er udvidet med 'NOT (facet.typeCategory="ebøger" OR facet.typeCategory="lydbøger")' for netop at ramme alt det andet der ikke er med i en af de to kategorier.

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:open="http://oss.dbc.dk/ns/opensearch">
  <SOAP-ENV:Body>
    <open:searchRequest>
      <open:query>klima NOT (facet.typeCategory="ebøger" OR
facet.typeCategory="lydbøger")</open:query>
      <open:agency>100200</open:agency>
      <open:profile>test</open:profile>
      <open:start>1</open:start>
      <open:stepValue>10</open:stepValue>
    </open:searchRequest>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```