

Find TEI ”ressurser” i CLARIN-DK



Introduktion

CLARIN-DK-infrastrukturen indeholder et væld af digitalt materiale, digitale forskningsobjekter, som i CLARIN-regi kaldes ressourcer.

I denne tutorial gennemgår vi hvordan du kan bruge de metadata som ressourcerne er opmærkede med, til at finde og evt. downloade de ressourcer som måtte være relevante for dig. Vær opmærksom på at alle tekstressourcer kun foreligger i xml.

Bemærk at du pt. ikke kan lave fritekstsøgning (søge på indholdet i selve ressursen), kun i metadata. Hvis du er interesseret i korpussøgning, kan du gå korpusplatformen Korp:

http://alf.hum.ku.dk/korp/#?stats_reduce=word&corpus=lsconstructioneb1,lsconstructioneb2&cqp=%5B%5D

Trin 1

Gå ind på: <https://clarin.dk/clarindk/find.jsp>

fx ved at klikke på "TEI-tekster i CLARIN-DK" på infrastrukturens forside: <https://clarin.dk/clarindk/forside.jsp>

Søg resurser i clarin.dk-arkivet

Metadata Avanceret tekstsøgning

Vælg ressurcetype

Tekst

Tekst

Media

Lyd

Video

Media-annotering

Tilføj Alle

Filtrer på rettigheder for indholdsdata

Public Academic Restricted

Metadatasøgning

Indtast søgeord

Avanceret Metadatasøgning

Vælg ressurcetype(r) = Indtast søgeord

Tilføj søgefelt

Find

Vælg ressurcetype

Tekst:

Tekstfiler som er konverteret til xml. Den oprindelige fil, fx Word eller PDF, er ikke længere tilgængelig. Se eksempel på TEI P5 xml-formatet i Appendiks 1.

Lingvistisk annotation af tekstfiler, fx tokenisering, lemmatisering eller POS-tagging. Annotationen ligger i en separat xml-fil adskilt fra teksten da der kan være forskellige ophavsrettigheder til de to. Se eksempel på annotationslag i Appendiks 1.

Filtrer på rettigheder for indholdsdata

Du vælger hvilke rettigheder de fremsøgte filer må have.

Public:

Resurser af denne type er tilgængelige for alle, også til kommerciel brug.

Academic:

Resurser af denne type er kun tilgængelige til akademisk brug. Den akademiske organisation man tilhører, skal være tilknyttet WAYF <http://www.wayf.dk/>

Restricted:

Resurser af denne type er kun tilgængelige for en begrænset gruppe defineret af dataleverandøren.

Metadatasøgning

Her kan du indtaste de(t) ord du ønsker at søge på fra felterne *title*, *creator*, *subject* og *description*. Søgemaskinen er ikke sensitiv over for store og små bogstaver men derimod over for begyndelsen af ordet. Du kan således søge på *Kutter Nyt* eller *kutter nyt* og få samme søgeresultat; men ikke på *utter Nyt*, **utter Nyt* eller *Kutter *yt*.

Det er selvsagt vanskeligt at vide hvad man skal søge efter, når man ikke ved hvad der findes i repositoryet. Vi har derfor lavet listen over resurser pr. 10/8 0216 <https://info.clarin.dk/clarin-dk-infrastrukturen/brug-clarindk/indhold/>, hvorfra man kan lade sig inspirere.

Eksempler på søgeord og hits:

- 1) Resursetype:** *tekst*
Metadatasøgning: *pesten*
Søgeresultat: 2 hits, 1 med titlen:
En kort Underretning Om Præservation og Curation imod Pesten: Efter Hans Kongel. Majestets, allernaadigeste Befaling ...
- 2) Resursetype:** *tekst*
Metadatasøgning: *helbred*
Søgeresultat: 4625 hits. Nogle med emnet *helbred (DK5-61)*, andre hvor *helbred* indgår i titlen fx *Skiftarbejde og helbred*

Avanceret metadatasøgning

Her kan du søge i specifikke metadatafelter ved at vælge et metadatafelt og angive den værdi du vil søge på.

Bemærk at du kan bruge operatorerne: = , <>, any, all.

= " indeholder"

!= " Indeholder ikke"

Any: finder ét af følgende ord

All: finder alle ord i tilfældig rækkefølge

Du kan også søge på flere felter samtidig ved at klikke på "Tilføj søgefelt". Bemærk at du her kan bruge de boolske søgeoperatorer AND, OR eller NOT dvs. finde fællesmængde, foreningsmængde eller differensmængde.

Eksempler på søgeord og hits:

- 1) Resursetype:** *tekst*
Avanceret metadatasøgning: *Title = helbred**
Søgeresultat: 22 hits hvor *helbred* indgår i titlen, enten som enkeltstående ord eller som en del af en sammensætning fx *Arbejdsmedicinsk helbredsattest*, *basisvejledning*
- 2) Resursetype:** *tekst*
Avanceret metadatasøgning: *Title = helbred*

Søgeresultat: 12 hits hvor *helbred* indgår i titlen som enkeltstående ord fx *Snedker, arbejde og helbred*

3) Resursetype: tekst

Avanceret metadatasøgning: Subject = *helbred (DK5-61)*, Title = *hjerte**

Søgeresultat: 50 hits med emnet *helbred (DK5-61)* og *hjerte* som en del af titlen. Fx *Hjerte-lunge-redning, basal, voksne*

4) Resursetype: tekst

Avanceret metadatasøgning: Subject <> *helbred (DK5-61)*, Title = *hjerte**

Søgeresultat: 11 hits hvor *hjerte* indgår i titlen, og hvor emnet ikke er *helbred (DK5-61)*. Fx *CTB version: Hjerte rimer på smerte.*

5) Resursetype: tekst

Avanceret metadatasøgning: Subject = *helbred (DK5-61)*, Title = *hjerte**, Publisher = *sundhed.dk*

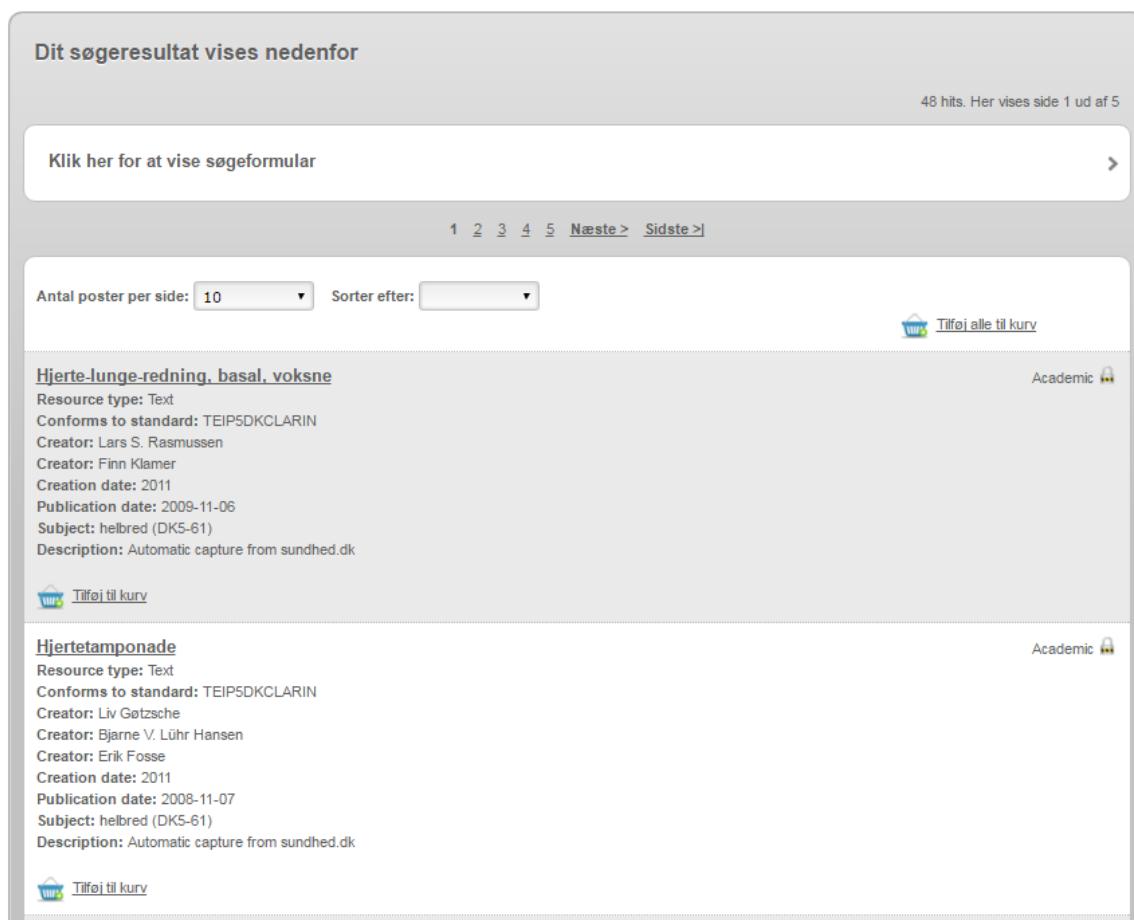
Søgeresultat: 48 hits med emnet *helbred (DK5-61)*, *hjerte* som en del af titlen og *sundhed.dk* som tekstleverandør.

Trin 2

Når du har indtastet dine søgekriterier og klikket *Find*, får du en liste af hits der matcher din søgning. Var søgningen for bred, kan du få uoverskueligt mange hits og bør derfor prøve at indsnævre søgningen. Få søgeformularen frem igen ved at klikke på bjælken *Klik her for at vise søgeformular*.



The header banner for CLARIN-DK features the logo on the left, a navigation menu in the center, and a search bar on the right. The navigation menu includes links for 'Forside', 'Introduktion', 'Overblik', 'Søg resurser', 'Søg i tekst', 'Deponer', 'Værktøjskasse', and 'Seneste søgning'. The search bar contains the text 'Resource ID' and an 'Ok' button. Below the navigation menu are buttons for 'Om Clarin', 'English', 'Gæst | Log ind', and 'Vis kurv 0' with a shopping cart icon.



The search results page displays a list of search results. At the top, it says 'Dit søgeresultat vises nedenfor' and '48 hits. Her vises side 1 ud af 5'. Below this is a button that says 'Klik her for at vise søgeformular' with a right-pointing arrow. A pagination bar shows '1 2 3 4 5 Næste > Sidste >'. Below the pagination bar are two dropdown menus: 'Antal poster per side: 10' and 'Sorter efter:'. To the right of these menus is a button that says 'Tilføj alle til kurv' with a shopping cart icon. The first search result is titled 'Hjerte-lunge-redning, basal, voksne' and is marked as 'Academic'. It includes details such as 'Resource type: Text', 'Conforms to standard: TEIP5DKCLARIN', 'Creator: Lars S. Rasmussen', 'Creator: Finn Klamer', 'Creation date: 2011', 'Publication date: 2009-11-06', 'Subject: helbred (DK5-61)', and 'Description: Automatic capture from sundhed.dk'. Below the first result is a button that says 'Tilføj til kurv' with a shopping cart icon. The second search result is titled 'Hjertetamponade' and is also marked as 'Academic'. It includes details such as 'Resource type: Text', 'Conforms to standard: TEIP5DKCLARIN', 'Creator: Liv Göttsche', 'Creator: Bjarne V. Lühr Hansen', 'Creator: Erik Fosse', 'Creation date: 2011', 'Publication date: 2008-11-07', 'Subject: helbred (DK5-61)', and 'Description: Automatic capture from sundhed.dk'. Below the second result is a button that says 'Tilføj til kurv' with a shopping cart icon.

Du kan sortere dit søgeresultat, lægge resurserne i kurven én ad gangen eller alle på siden. Du kan også vælge at se nærmere på de fremfundne resurser ved at klikke på titlen.

Hjerte-lunge-redning, basal, voksne

[Detaljer](#)
[CMDI metadata](#)
[Deponerede metadata](#)
[Relationer](#)
[Download](#)
[Vis indholdsdata](#)
[Tilføj til kurv](#)

Title: Hjerte-lunge-redning, basal, voksne
Language: da
Resource type: Text
Creator: Lars S. Rasmussen
Creator: Finn Klamer
Creation date: 2011
Subject: helbred (DK5-61)
Description: Automatic capture from sundhed.dk

Source title: Hjerte-lunge-redning, basal, voksne
Publisher: sundhed.dk
Publication date: 2009-11-06

Sponsor: DK-CLARIN
Data provider: cst.ku.dk
Data provider: dsn.dk

Format: text/xml
Conforms to standard: TEIP5DKCLARIN

Size: 570 words
Size: 63 paragraphs

Når du har klikket på en resurse, ser du 6 faneblade:

- Det første **Detaljer** viser lidt flere metadata end oversigtslisten.
- Under **CMDI metadata** ser du ressursens metadata konverteret til CMDI, som er udvekslingsformatet i det paneuropæiske CLARIN.
- **Deponerede metadata** svarer til CMDI-metadataene, blot er de i TEI P5 som er det tekst- og metadataformat der kræves i CLARIN-DK.
- Under **Relationer** kan du se om den valgte resurse har fx lingvistiske annotationer knyttet til sig.
- Hvis du er logget ind, kan du downloade ressursen med annotationer under **Download**. Ved at klikke på *Tilføj til kurv* har du dog flere muligheder ift. downloadformatet.
- Endelig kan du se selve indholdet af ressursen under **Vis indhold**.

Teksten ses uden den oprindelige formatering, én sætning pr. linje. Når der er billeder tilknyttet tekstresser, fx af originalteksten, kan de ses til venstre og teksten til højre.

Trin 3

- Når du har fundet de ressourcer du vil downloade, lægger du dem i kurven (én ad gangen eller alle på en side) og går derefter til kurveikonet *Vis kurv*. Fra kurven kan du enten vælge *Værktøjskasse* (se [Værktøjskassen i Clarin.dk: Lingvistisk opmærkning](#)) eller *Download*.

Download

Download af flere filer

Systemet har mulighed for at pakke de ressourcer der ligger i din kurv på forskellige måder, vælg en af metoderne nedenfor. Hvis du forsøger at downloade ressourcer du ikke har de nødvendige rettigheder til at downloade, vil du blot hente ressourcens metadata, og ikke indholdsdata. Husk evt. at logge ind, da det kan give adgang til flere ressourcers indholdsdata.

Vælg metode

- Download de valgte ressourcer med metadata og indholdsdata**
De valgte ressourcer med data og metadata vil blive pakket i en zip-fil.
- Download de valgte ressourcer med metadata og indholdsdata, og inkluder også eventuelle annotationer**
De valgte ressourcer og deres annotationer vil blive pakket i en zip-fil.
- Flet de valgte tekster sammen med deres annotationer i et TEIP5-dokument.**
Dette virker kun på tekster og annotationer i TEIP5DKCLARIN format! Du skal kun lægge teksterne i kurven, systemet finder selv annotationerne via relationerne. Resultatet af sammenfletningen vil blive pakket i en zip-fil.
- Flet de valgte tekster sammen med deres annotationer i CQP-format.**
Dette virker kun på tekster og annotationer i TEIP5DKCLARIN format! Du skal kun lægge teksterne i kurven, systemet finder selv annotationerne via relationerne. CQP-formatet kan bruges i [IMS Open Corpus Workbench](#). Resultatet af sammenfletningen vil blive pakket i en zip-fil.

E-mail

Når arbejdet er udført, vil du modtage en e-mail med link til zip-filen som du kan downloade. Hvis der ikke er anført en e-mail-adresse i feltet herunder, skal du tilføje en e-mail-adresse.

E-mail

Næste

Før du kommer til *Download-siden* kan du blive bedt om at logge ind på clarin.dk. Dette sker automatisk gennem WAYF og den adgang du har til din (akademiske) institution¹. Filerne du downloader, kan pakkes og formateres på forskellig vis alt efter om du har brug for at se:

- annotationerne isoleret
- har brug for at annotationerne er flettet sammen med teksten i xml TEI P5-format eller
- CQP-format til korpussøgning.

Eksempler:

1. Du er kun interesseret i indholdsdata og metadata.

Vælg *Download de valgte ressourcer med metadata og indholdsdata*

Du får tilsendt et link til en zip-fil der indeholder en række filer. Foruden den originale deponerede fil *original-deposited-file*, der er i TEI P5-format for tekstfiler, indeholder zip-filen: *dkclarin-NR-cmdi.xml* og *dkclarin-NR-teiheader.xml*, som indeholder metadata i TEI P5 og i CMDI-format, samt *filnavn-NR-unite.xml* der foruden indholdsdata også indeholder annotationer hvis de findes med samme licens som indholdsdata. Du kan se eksempel på indholdsdata og annotationer i TEI P5 i Appendiks 1.

¹ Når du skal logge ind gennem WAYF, så er det første gang en god ide at vælge "Danmark" i feltet "Show providers in" eller at søge efter din institution i søgefeltet.

2. Du er interesseret i indholdsdata, metadata og annotationer.

Vælg *Download de valgte ressourcer med metadata og indholdsdata, og inkluder også eventuelle annotationer*

Du får tilsendt et link til en zip-fil indeholdende en række filer som beskrevet ovenfor. Desuden er eventuelle annotationer pakket i separate direktorier hvor annotationsfilen fx kan hedde *filnavn-NR-unite-cstlemma.xml* eller *filnavn-NR-unite-CstClarinDaTokeniser.xml* alt efter om det er lemma- eller tokeniseringsannotation der er tale om.

Du kan se et eksempel på indholdsdata i TEI P5-format i Appendiks 1

3. Du er interesseret i at annotationerne er flettet sammen med indholdsdata i TEI P5-format

Vælg *Flet de valgte tekster sammen med deres annotationer i et TEIP5-dokument.*

Dette virker kun på tekster og annotationer i TEIP5DKCLARIN-format! Du skal kun lægge teksterne i kurven, systemet finder selv annotationerne via relationerne. Du får tilsendt et link til en zip-fil som er resultatet af sammenfletningen. Foruden sammenfletningen af indholdsdata og annotationer i filen *dkclarin-NR-merge.xml*, indeholder zip-filen også en række filer med metadata.

Se eksempel på det sammenflettede TEI P5-format i Appendiks 1

4. Du er interesseret i at annotationerne er flettet sammen med indholdsdata i CQP-format

Vælg *Flet de valgte tekster sammen med deres annotationer i CQP-format.*

Dette virker kun på tekster og annotationer i TEIP5DKCLARIN-format! Du skal kun lægge teksterne i kurven, systemet finder selv annotationerne via relationerne. CQP-formatet kan bruges i IMS Open Corpus Workbench.

Du får tilsendt et link til en zip-fil som er resultatet af sammenfletningen. Foruden sammenfletningen af indholdsdata og annotationer i filen *dkclarin-NR-merge.xml*, indeholder zip-filen også en række filer med metadata.

Se eksempel på det sammenflettede CQP-format i Appendiks 1

Appendiks 1 viser eksempler på xml-formatet TEI P5, som er det tekstformat CLARIN-DK-infrastrukturen anvender. Først ses indholdsdata og annotationer i TEI P5, derefter et downloadformat hvor annotationer og indholdsdata er flettet sammen og sidst CQP-downloadformatet.

I CLARIN-DK er der pr. 10/8 2016 følgende lingvistiske annotationstyper på en række tekster:

- Token
- POS-tag
- Lemma
- Termvægt

1. Eksempel på indholdsdata og annotationer i en TEI P5 xml-fil

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:schemaLocation="http://www.tei-c.org/ns/1.0
http://dkclarin.dk/schemas/WP2/TEIDKCLARINv1.0/xml.xsd">
<teiHeader type="text">
....
  </teiHeader>
<text>
  <body>
<p n="1">
    <w xml:id="i1.1">Region</w>
    <c xml:id="i1.2" type="s"/>
    <w xml:id="i1.3">Hovedstaden</w>
    <c xml:id="i1.4" type="s"/>
    <w xml:id="i1.5">Specialebeskrivelser</w>
</p>
<p n="2">
    <w xml:id="i2.1">Klinisk</w>
    <c xml:id="i2.2" type="s"/>
    <w xml:id="i2.3">Immunologi</w>
</p>
<p n="3">
    <w xml:id="i3.1">Februar</w>
    <c xml:id="i3.2" type="s"/>
    <w xml:id="i3.3">2006</w>
</p>
<p n="4">
    <w xml:id="i4.1">Specialebeskrivelse</w>
    <c xml:id="i4.2" type="p">:</c>
    <c xml:id="i4.3" type="s"/>
    <w xml:id="i4.4">Klinisk</w>
    <c xml:id="i4.5" type="s"/>
    <w xml:id="i4.6">Immunologi</w>
....
</p>
</body>
<spanGrp ana="#CstClarindaTokeniser">
  <span xml:id="t1" from="i1.1">Region</span>
  <span xml:id="t2" from="i1.3">Hovedstaden</span>
  <span xml:id="t3" from="i1.5">Specialebeskrivelser</span>

```

```

    <span xml:id="t4" from="i2.1">Klinisk</span>
    <span xml:id="t5" from="i2.3">Immunologi</span>
    <span xml:id="t6" from="i3.1">Februar</span>
    <span xml:id="t7" from="i3.3">2006</span>
    <span xml:id="t8" from="i4.1">Specialebeskrivelse</span>
    <span xml:id="t9" from="i4.2">:</span>
    <span xml:id="t10" from="i4.4">Klinisk</span>
    <span xml:id="t11" from="i4.6">Immunologi</span>
    ....
</spanGrp>
<spanGrp ana="#CstClarinSentenceAndParagraphSegmenter">
    <span xml:id="S1" type="Sseg" from="t1" to="t3"/>
    <span xml:id="P1" type="Pseg" from="t1" to="t3"/>
    <span xml:id="S2" type="Sseg" from="t4" to="t5"/>
    <span xml:id="P2" type="Pseg" from="t4" to="t5"/>
    <span xml:id="S3" type="Sseg" from="t6" to="t7"/>
    <span xml:id="P3" type="Pseg" from="t6" to="t7"/>
    <span xml:id="S4" type="Sseg" from="t8" to="t11"/>
    <span xml:id="P4" type="Pseg" from="t8" to="t11"/>
    ....
</spanGrp>
<spanGrp ana="#csttaggerXML">
    <span xml:id="p1" from="t1">N_INDEF_SING</span>
    <span xml:id="p2" from="t2">N_DEF_SING</span>
    <span xml:id="p3" from="t3">V_PRESENT</span>
    <span xml:id="p4" from="t4">ADJ</span>
    <span xml:id="p5" from="t5">N_INDEF_SING</span>
    <span xml:id="p6" from="t6">N_INDEF_SING</span>
    <span xml:id="p7" from="t7">NUM</span>
    <span xml:id="p8" from="t8">N_INDEF_SING</span>
    <span xml:id="p9" from="t9">TEGN</span>
    <span xml:id="p10" from="t10">ADJ</span>
    <span xml:id="p11" from="t11">N_INDEF_SING</span>
    ....
</spanGrp>
<spanGrp ana="#Cstlemma">
    <span xml:id="l1" from="t1" >region</span>
    <span xml:id="l2" from="t2" >hovedstad</span>
    <span xml:id="l3" from="t3" >Specialebeskrivelser</span>
    <span xml:id="l4" from="t4" >klinisk</span>
    <span xml:id="l5" from="t5" >immunologi</span>
    <span xml:id="l6" from="t6" >februar</span>
    <span xml:id="l7" from="t7" >2006</span>
    <span xml:id="l8" from="t8" >Specialebeskrivelse</span>
    <span xml:id="l9" from="t9" >:</span>
    <span xml:id="l10" from="t10" >klinisk</span>
    <span xml:id="l11" from="t11" >immunologi</span>
    ....
</spanGrp>
</text>
</TEI>

```

2. Eksempel på annotationer flettet sammen med indholdsdata i TEI P5-format:

```
<p n="4">
  <phr type="paragraph" xml:id="P4">
    <s type="sentence" xml:id="S4">
      <seg xml:id="seg8">
        <spanGrp>
          <span xml:id="t8" type="token" ana="#CstClarinDaTokenizer">Dykning</span>
          <span xml:id="l8" type="lemma" ana="#cstlemma">Dykning</span>
          <span xml:id="th8" type="term weight" ana="#DsnClarinDaTermTaggerLogLikelihood">152.891</span>
          <span xml:id="p8" type="POS-tag" subtype="CSTtags" ana="#csttaggerXML">N_INDEF_SING</span>
        </spanGrp>
        <w xml:id="i4.1">Dykning</w>
      </seg>
      <c xml:id="i4.2" type="s"/>
      <seg xml:id="seg9">
        <spanGrp>
          <span xml:id="t9" type="token" ana="#CstClarinDaTokenizer">foregår</span>
          <span xml:id="l9" type="lemma" ana="#cstlemma">foregå</span>
          <span xml:id="th9" type="term weight" ana="#DsnClarinDaTermTaggerLogLikelihood">4.924</span>
          <span xml:id="p9" type="POS-tag" subtype="CSTtags" ana="#csttaggerXML">V_PRESENT</span>
        </spanGrp>
        <w xml:id="i4.3">foregår</w>
      </seg>
      <c xml:id="i4.4" type="s"/>
      <seg xml:id="seg10">
        <spanGrp>
          <span xml:id="t10" type="token" ana="#CstClarinDaTokenizer">på</span>
          <span xml:id="l10" type="lemma" ana="#cstlemma">på</span>
          <span xml:id="th10" type="term weight" ana="#DsnClarinDaTermTaggerLogLikelihood">0</span>
          <span xml:id="p10" type="POS-tag" subtype="CSTtags" ana="#csttaggerXML">PRÆP</span>
        </spanGrp>
        <w xml:id="i4.5">på</w>
      </seg>
      <c xml:id="i4.6" type="s"/>
      <seg xml:id="seg11">
        <spanGrp>
          <span xml:id="t11" type="token" ana="#CstClarinDaTokenizer">mange</span>
          <span xml:id="l11" type="lemma" ana="#cstlemma">mangen</span>
          <span xml:id="th11" type="term weight" ana="#DsnClarinDaTermTaggerLogLikelihood">-0.066</span>
          <span xml:id="p11" type="POS-tag" subtype="CSTtags" ana="#csttaggerXML">ADJ</span>
        </spanGrp>
        <w xml:id="i4.7">mange</w>
      </seg>
      <c xml:id="i4.8" type="s"/>
      <seg xml:id="seg12">
        <spanGrp>
          <span xml:id="t12" type="token" ana="#CstClarinDaTokenizer">forskellige</span>
          <span xml:id="l12" type="lemma" ana="#cstlemma">forskellig</span>
          <span xml:id="th12" type="term weight" ana="#DsnClarinDaTermTaggerLogLikelihood">19.936</span>
          <span xml:id="p12" type="POS-tag" subtype="CSTtags" ana="#csttaggerXML">ADJ</span>
        </spanGrp>
        <w xml:id="i4.9">forskellige</w>
      </seg>
      <c xml:id="i4.10" type="s"/>
      <seg xml:id="seg13">
        <spanGrp>
          <span xml:id="t13" type="token" ana="#CstClarinDaTokenizer">måder</span>
          <span xml:id="l13" type="lemma" ana="#cstlemma">måde</span>
```

7.050
N_INDEF_PLU
</spanGrp>
<w xml:id="i4.11">måder</w>
</seg>

3. Eksempel på annotationer flettet sammen med indholdsdata i CQP-format

```

<p>
<s>
Dykning      Dykning      152.891      N_INDEF_SING
foregår      foregå       4.924        V_PRESENT
på           på           0            PRÆP
mange       mangen      -0.066       ADJ
forskellige  forskellig  19.936       ADJ
måder       måde        7.050        N_INDEF_PLU
og          og          0            SKONJ
niveauer    niveau      12.332       N_INDEF_PLU
(           (           0            TEGN
dykkerklokke  dykkerklokke 27.740      N_INDEF_SING
,           ,           0            TEGN
selvforsynet  selvforsyn  28.557       N_DEF_SING
,           ,           0            TEGN
slange      slange      14.823       N_INDEF_SING
,           ,           0            TEGN
fritid      fritid      9.452        N_INDEF_SING
,           ,           0            TEGN
professionelt  professionel 10.925       ADJ
)           )           0            TEGN
og          og          0            SKONJ
dette      denne      13.162       PRON_DEMO
er         være       0            V_PRESENT
en         en         0            PRON_UBST
kort       kort       2.605        ADJ
indføring  indføring  16.839       N_INDEF_SING
i         i         0            PRÆP
helbredsaspekter  helbredsaspekt 31.600      N_INDEF_PLU
ved       ved       17.189       PRÆP
dykning   dykning   152.891      N_INDEF_SING
</s>
</p>

```