

Minimal use case

Suggestions from Lene Offersgaard, CLARIN-DK

Goal : to be able to combine more than one DC metadata information with direct text content search. These kind of searches are useful to end-users looking for relevant resources to extract.

Fx: To be able to search for a set of texts that:

A: are **from 1980'ies**

with the named entity 'Vesterbro' (a part of Copenhagen) in the **description** or in text **content**

B: are **from 2000-2009**

with **subject "health"**
content word "stress"

Next steps

Goal: to be able to combine advanced metadata search: search on metadata content that are out-of-DC-domain, combined with more than one restriction on content words (lemma search).

Result set: resource set (if downloading resources are allowed) or extracted sentences

Search for resources covering e.g. "expert-expert" or "expert-layman" communication, for subject domain "health" with content lemma for "stress" and "blodtryk"(blood pressure).

Here users can compare the texts written from expert to expert, with expert to layman.

Adding lemma search: then we need to agree on how to search in a lemma annotation across repositories

Difficult

Pos search seems to be complicated:

- as the same text might have more than one pos-annotation!
- pos-tags are named differently across taggers

We could add a special pos-annotation that translates from proprietary formats to common pos-morphology.