# Topic: search engine upgrades: manual annotation

**IDs in Google Sheets:** T85

**Author:** INT

**What exists:** static corpora and search engines; annotation tools

**What must be adapted / extended / created anew:** Create an infrastructure for managing user annotations in corpora. All corpus search applications should support the manual annotation of search results with codes to further refine/classify the search results (cf. the Lancaster BNC interface).

* Users should be able to add annotations to search results and to share these annotations with other users.
* Deposit (manual) annotation results back into an archive (closing the data management life cycle)
* Annotations should be searchable in the corpus (depends on design approach)
* Should also benefit treebank search

There are two basic scenarios: either there is a strong integration with the search engine, which would simplify support of making annotations searchable; or this is a module that can be combined with any type of search engine.

Example use cases:

* Bookmarking function
* Annotate word senses
* Annotate tagger errors
* Add fine-grained syntactic information

Cf:

* bnc web annotated queries (screenshots below)
* <https://github.com/emanjavacas/cosycat/wiki/Annotations>

**Why important for CLARIAH (scientific impact):** Added value to research with corpora.

**Targeted/Actual users:** corpus users

**Actual use (quantify!):** ##

**Social Impact** **(concrete examples):** Facilitate linguistic research; publish data on which research is based; foster collaborative corpus research

**Estimate in PMs (try to justify):** 6PM

**Lead + PMs:** XX:

**Participants + PMs:** INT/UU

**Working plan; deliverables**

In year 1, requirements will be analyzed and we will draft a technical design document. The main challenge is to devise a way to support different search engine platforms with minimal adaptations to the underlying corpus retrieval systems.

in year 2, backend functionality will be developed

It is not unlikely that the annotation environment will be based on the Lex’it platform that INT uses internally for data production.

Totaal 6PM

Year 1: analysis of requirements and design of technical architecture; UI requirements

[total 2 PM; developer (Jan), CL (Carole, Vincent), Linguists]

Year 2: implementation: blacklab support for search of annotations

[2 PM backend developer]

[2 PM frontend developer]

Milestone: year 1: document describing requirements and technical design

Mlinestone: year 2: backend implementation

Deliverable: year 3: autosearch, blacklab and blacklab server versions with handling of manual annotation





