# T85: search engine upgrades: manual annotation

**IDs in Google Sheets:** T85

**Author:** INT/Jesse de Does

**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 one implements a module that can be combined with any type of search engine.  
Example use cases:

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

Cf also:

* 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:**

**Participants + PMs: INT**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Taak** | **Type personeel** | **Available** | **2019** | **2020** | **2021** | **2022** | **2023** | **Summed** | **Task total** |
| T085 Manual Annotation | Software developer(s) | 5 | 1 | 2 | 2 |  |  | **5** | **6** |
| T085 Manual annotation | (Computational) linguist | 1 | 1 |  |  |  |  | **1** |  |

**Deliverables:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | ID | D or M | Type | Description |
| T085 Manual Annotation | T085M1 | Milestone | Document | Document defining the requirements for the manual annotation of corpus search results, based on input from a set of use cases. |
|  | T085D2 | Deliverable | Software | Implementation manual search result annotation, based on BlackLab and the BlackLab corpus frontend |
|  | T005D2 | Deliverable | Document | Document describing T085D2 and its application to use cases. |