Semantic Web @ HUG

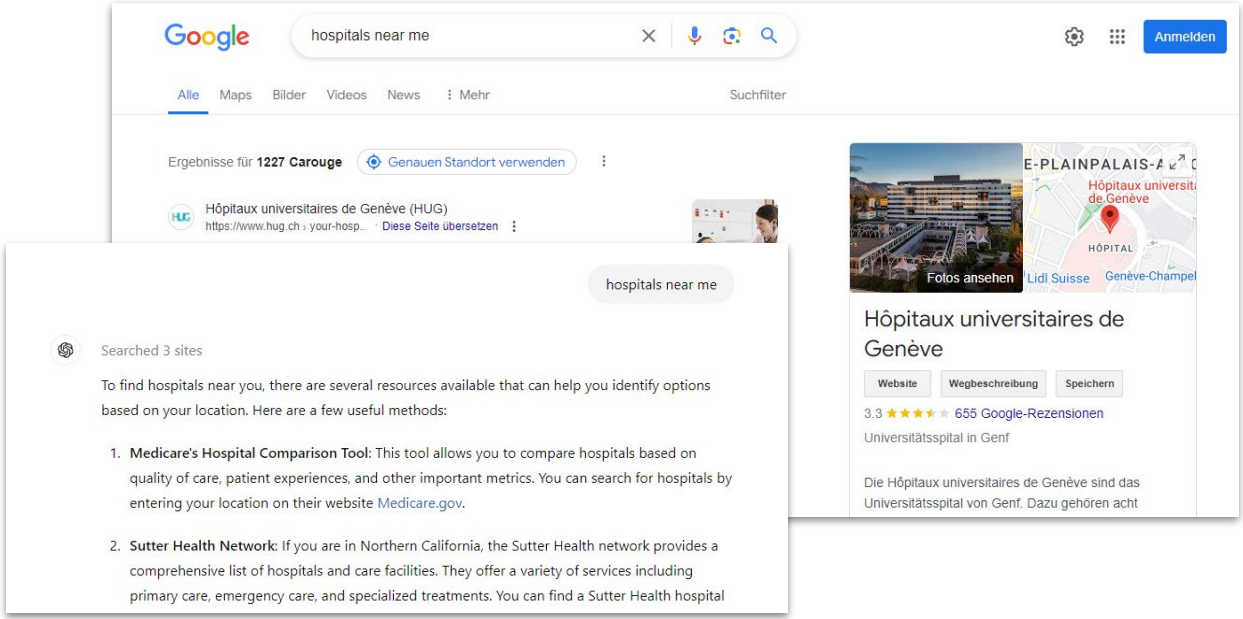
Viktor Shcherbakov

Problem statement

Hôpitaux Universitaires Genève

Evaluate the effect of semantic annotation on:

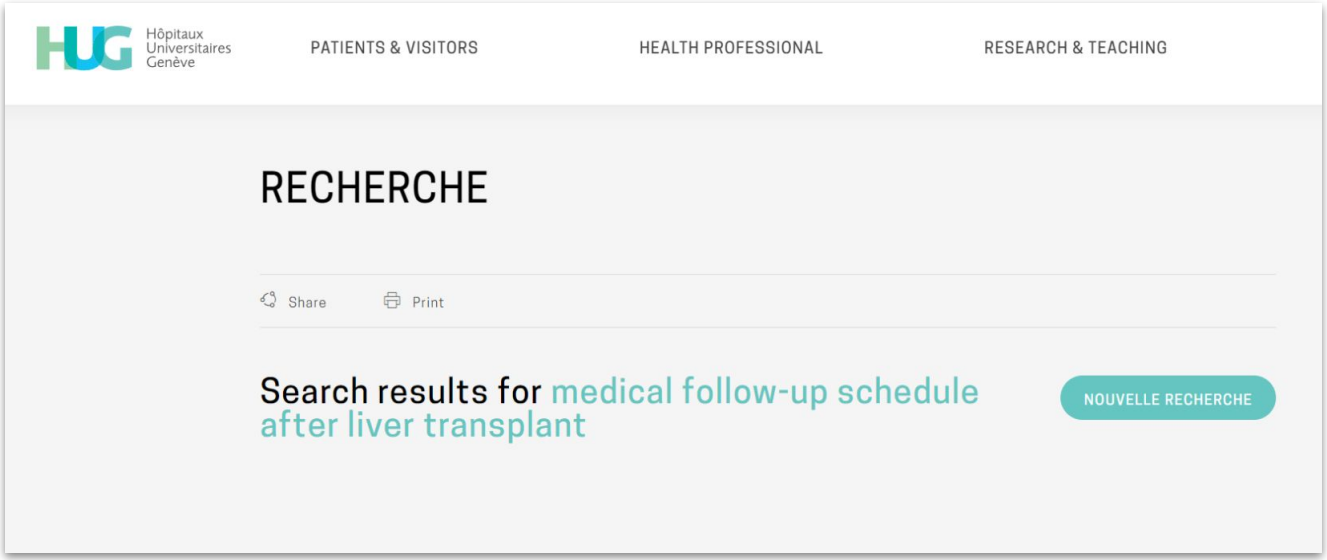
1. External search
2. Internal search
3. Chatbot search



Feasibility

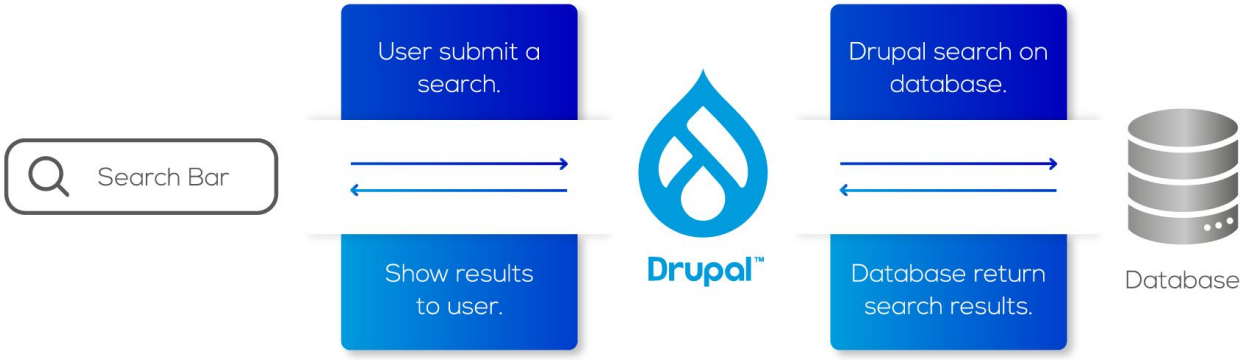
Evaluate the effect of semantic annotation on:

1. ~~External search~~
2. Internal search
3. ~~Chatbot search~~



Internal search | Drupal

Hôpitaux Universitaires Genève

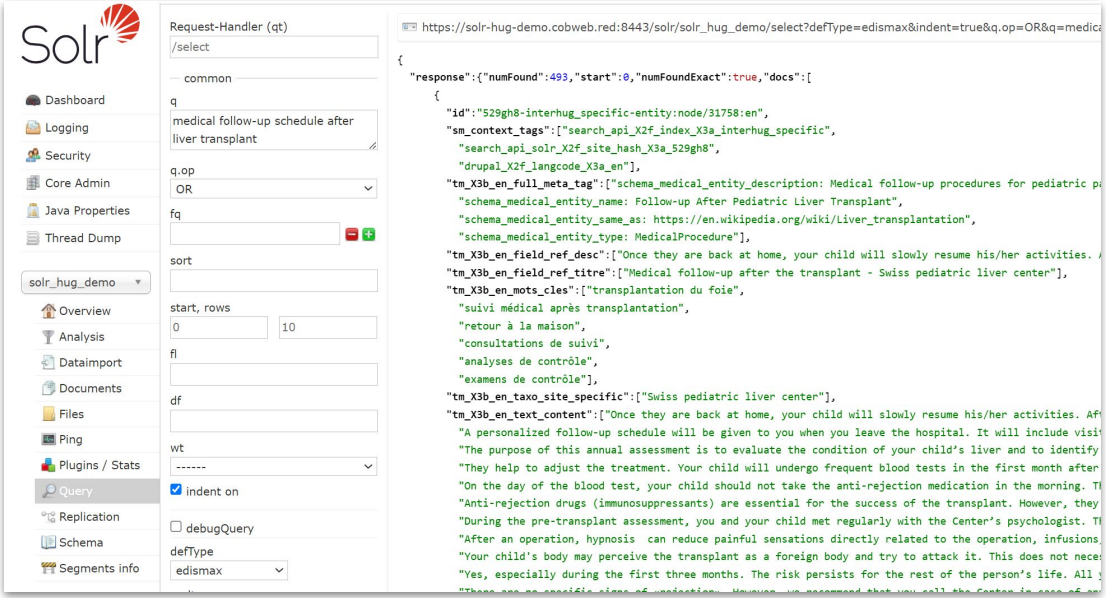


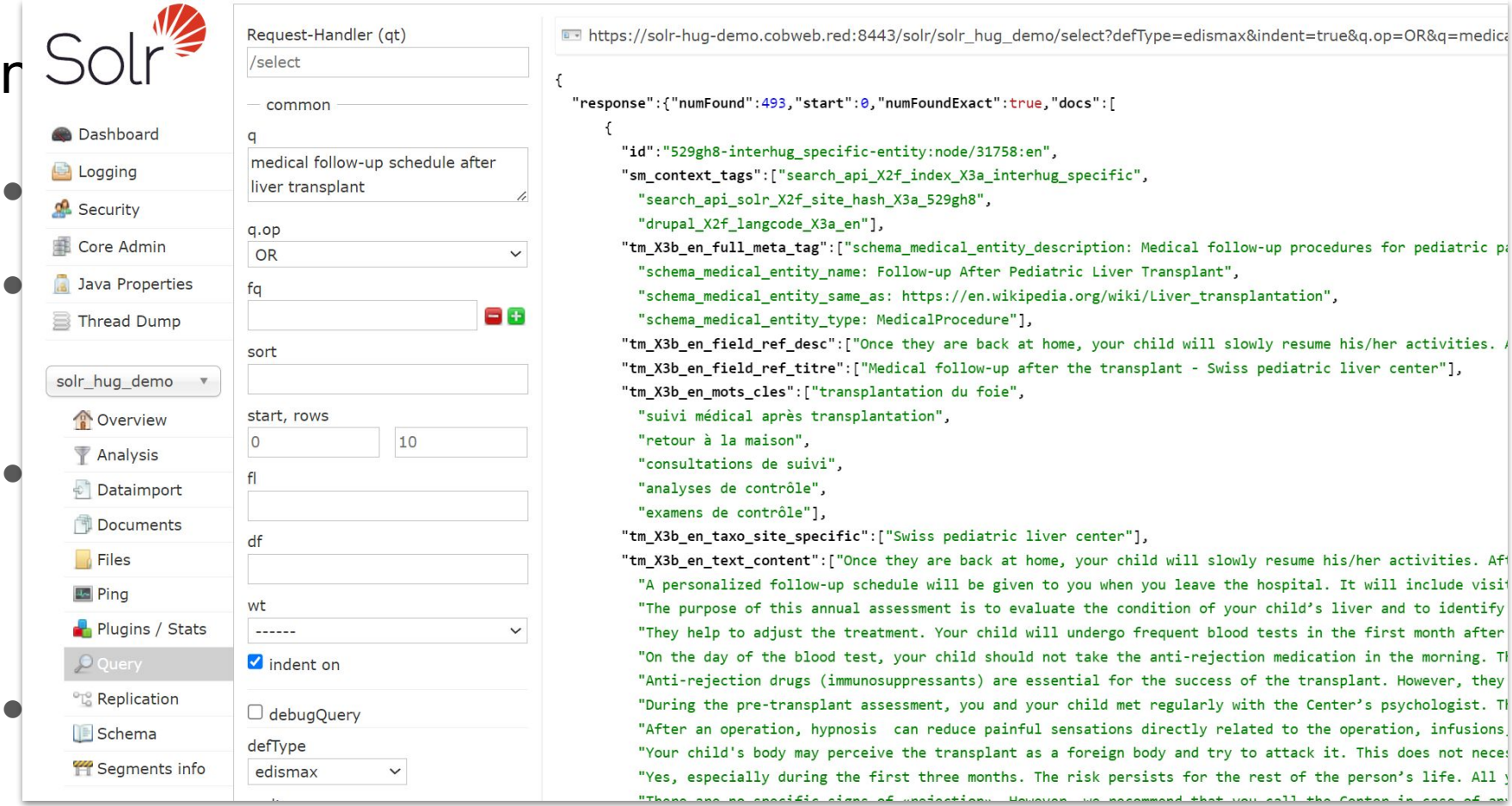
Defines

- how to index the content;
- how to display search results to user

Internal search | Solr

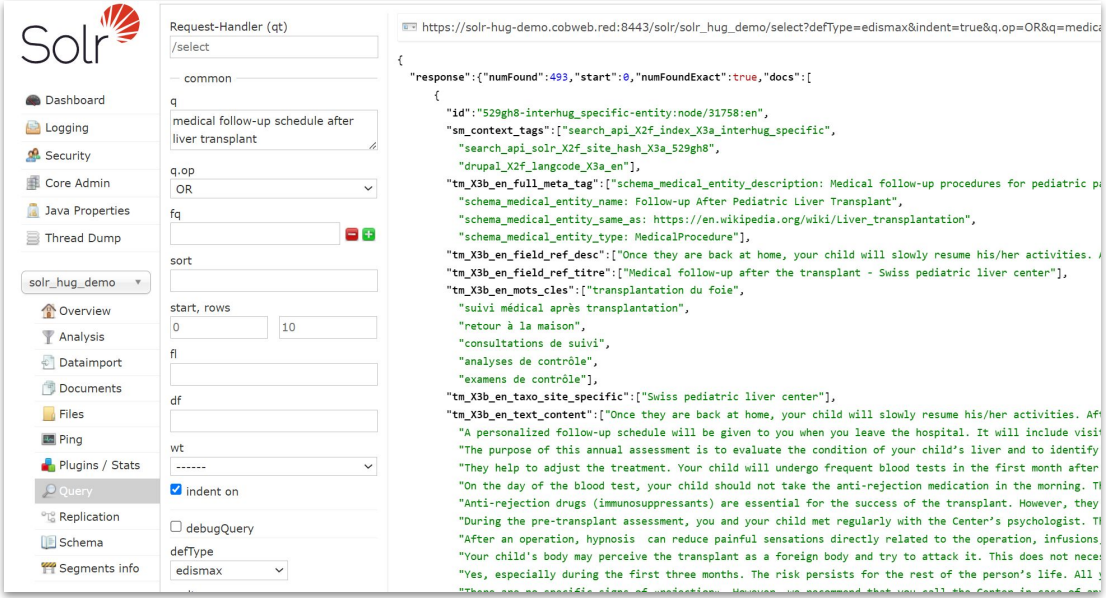
- Optimized for Full-Text search
- Without additional extensions works as term matching of queries and documents
- Includes many term extraction techniques like lemmatization, stemming, n-gram, synonym mapping, etc.
- BM25 as core information retrieval model




Solr
Dashboard
Logging
Security
Core Admin
Java Properties
Thread Dump
solr_hug_demo
Overview
Analysis
Dataimport
Documents
Files
Ping
Plugins / Stats
Query
Replication
Schema
Segments info
Request-Handler (qt)
/select
common
q
medical follow-up schedule after liver transplant
q.op
OR
fq
sort
start, rows
0
10
fl
df
wt
------
indent on
debugQuery
defType
edismax
https://solr-hug-demo.cobweb.red:8443/solr/solr_hug_demo/select?defType=edismax&indent=true&q.op=OR&q=medica
{
"response":{"numFound":493,"start":0,"numFoundExact":true,"docs":[
{
"id":"529gh8-interhug_specific-entity:node/31758:en",
"sm_context_tags":["search_api_X2f_index_X3a_interhug_specific",
"search_api_solr_X2f_site_hash_X3a_529gh8",
"drupal_X2f_langcode_X3a_en"],
"tm_X3b_en_full_meta_tag":["schema_medical_entity_description: Medical follow-up procedures for pediatric pa
"schema_medical_entity_name: Follow-up After Pediatric Liver Transplant",
"schema_medical_entity_same_as: https://en.wikipedia.org/wiki/Liver_transplantation",
"schema_medical_entity_type: MedicalProcedure"],
"tm_X3b_en_field_ref_desc":["Once they are back at home, your child will slowly resume his/her activities. A
"tm_X3b_en_field_ref_titre":["Medical follow-up after the transplant - Swiss pediatric liver center"],
"tm_X3b_en_mots_cles":["transplantation du foie",
"suivi médical après transplantation",
"retour à la maison",
"consultations de suivi",
"analyses de contrôle",
"examens de contrôle"],
"tm_X3b_en_taxo_site_specific":["Swiss pediatric liver center"],
"tm_X3b_en_text_content":["Once they are back at home, your child will slowly resume his/her activities. Af
"A personalized follow-up schedule will be given to you when you leave the hospital. It will include visi
"The purpose of this annual assessment is to evaluate the condition of your child's liver and to identify
"They help to adjust the treatment. Your child will undergo frequent blood tests in the first month after
"On the day of the blood test, your child should not take the anti-rejection medication in the morning. Th
"Anti-rejection drugs (immunosuppressants) are essential for the success of the transplant. However, they
"During the pre-transplant assessment, you and your child met regularly with the Center's psychologist. Th
"After an operation, hypnosis can reduce painful sensations directly related to the operation, infusions
"Your child's body may perceive the transplant as a foreign body and try to attack it. This does not neces
"Yes, especially during the first three months. The risk persists for the rest of the person's life. All y
"There are no specific signs of «rejection». However, we recommend that you call the Center in case of an

Internal search | Solr

- Optimized for Full-Text search
- Without additional extensions works as term matching of queries and documents
- Includes many term extraction techniques like lemmatization, stemming, n-gram, synonym mapping, etc.
- BM25 as core information retrieval model



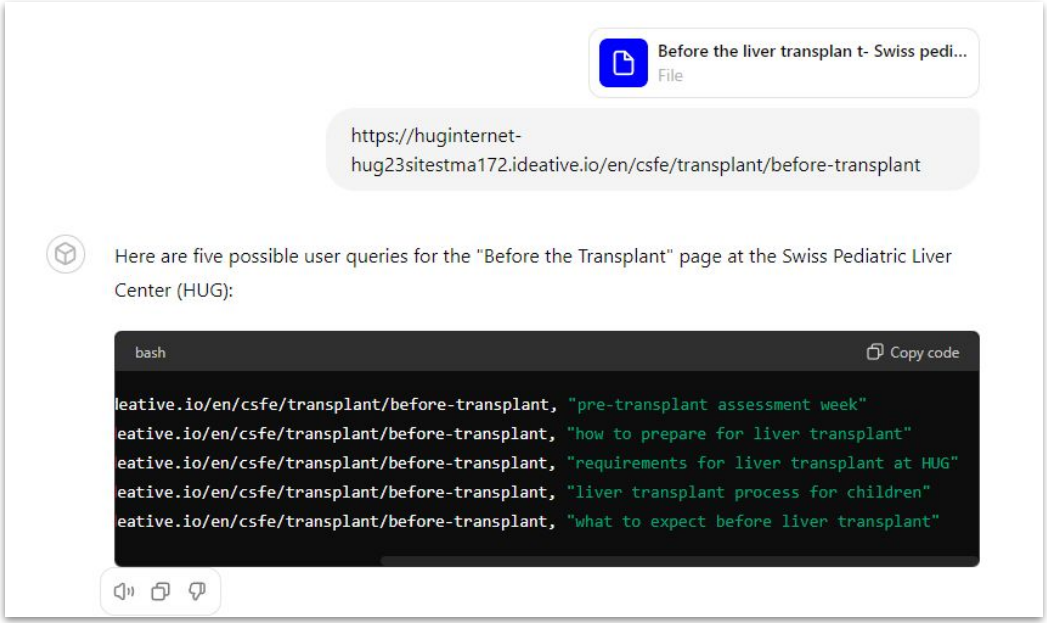
Internal search | BM25

$$\text{score}(D, Q) = \sum_{i=1}^{n} \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}$$

- Rare terms (across all documents) contribute more
- Frequent document terms contribute more
- Score can be influenced with free parameters controlling the contribution of document length and term frequency

Evaluation method | Dataset

- Focus on CSFE webpages
- Generation with LLM
- 247 queries
- 51 webpages



Link to GPT

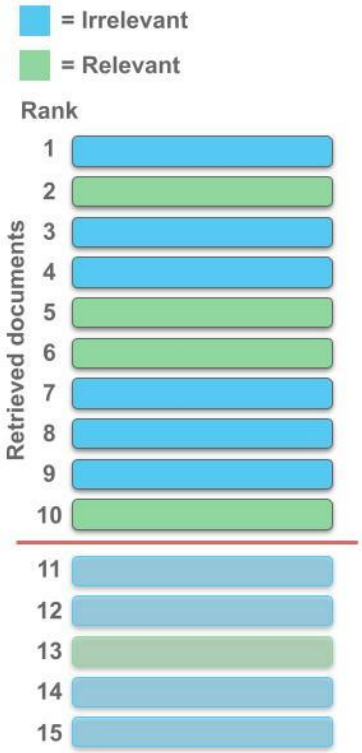
Link to before-transplant

Evaluation method | Metric

Hôpitaux Universitaires Genève

$$Recall@k = \frac{1}{N}\sum_{i=1}^{N}\frac{\#\text{ of retrieved relevant documents for query } i}{\#\text{ of relevant documents for query } i}$$

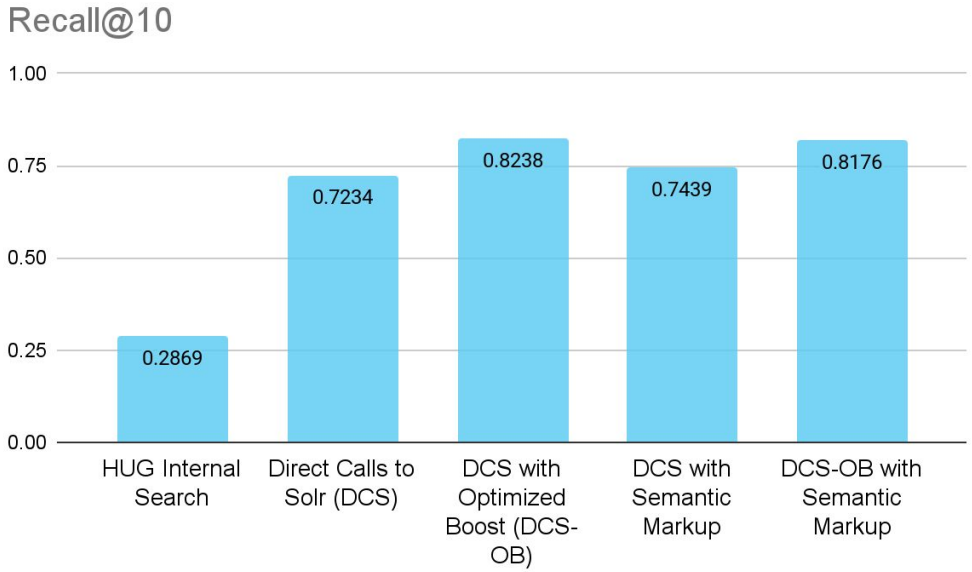
- Can be interpreted as a probability to retrieve **all** relevant documents at the first page of the search.
- The search system with high Recall at *k* is expected to retrieve **all** relevant documents and rank them in a way, that most of them end up on the **first** page



Evaluation results

Each search field in Solr has boost weights for each field.

Versions with Optimized Boost refer to Powell optimization of said weights



Case study

Hôpitaux Universitaires Genève

Try to find **https://www.hug.ch/en/csfe/transplant/before-transplant**

Example query: requirements for liver transplant at HUG

Google Search results

HUG Internal Search results

Link to before-transplant

Conclusions and Recommendations

Hôpitaux Universitaires Genève

- Current state of the internal search is suboptimal to say the least
- Fortunately, current system can be optimized to yield decent results
- A representative dataset of user queries needs to be collected
- Semantic markup affects the search marginally

Thank you!