Lab Project
Milestone IV: Research Assignment

Derive and test a research question with your IR system.

- ❑ Come up with a hypothesis that can be tested with your system

- ❑ Conduct statistical analysis to address the hypothesis
  - – Hyp. should separate two configurations (modified c., control c.)
  - – Hyp. should motivate a distribution over both groups to test against

- ❑ Summarize your system and findings in a short report

"System effectiveness will be higher when boosting
the rank of documents under 1000 characters."

- ❑ Due Date: 08.01.2024
- ❑ Deliverable: Report (max. 2 pages)

Lab Project
Milestone IV: Research Assignment

How is this different from Milestone III?

- ❏ M3: baseline system that accomplishes initial retrieval
    - System effectiveness is not the main focus
    - Intended to get comfortable with libraries & Tira

- ❏ M4: improved system that aims for best effectiveness
    - Hypothesis-driven development
    - Improved version of the M3 system; New stage on top of M3 system; ...

Lab Project
Milestone IV: Research Assignment

What makes a good research question?

Lab Project
Milestone IV: Research Assignment

What makes a good research question?

❑ A good research question... [Bartos 1992]

– ...investigates about the relationship between two or more variables.
– ...is testable (i.e., it is possible to collect data to answer the question).
– ...is stated as clearly and as non-trivial question (i.e., includes all relevant information, not answerable Yes/No).
– ...does not pose an ethical or moral problem for implementation.
– ...is specific and restricted in scope.
– ...identifies exactly what is to be solved.

Lab Project

What makes a good hypothesis?

Lab Project
Milestone IV: Research Assignment

What makes a good hypothesis?

- ❑ A good hypothesis. . .

    - – . . . is grounded in a research question.

    - – . . . is testable.

    - – . . . states an expected relationship between variables.

    - – . . . is stated as simply and concisely as possible.

- ❑ Hypothesis testing:

    - – Step 1: What are your variables? (nominal, ordinal, scale, ratio)

    - – Step 2: Measure the variables (Are aggregated measures enough?)

    - – Step 3: Significance test (Null hypothesis? Which α level? Which significance test?)

Lab Project

- ❑ Example 1 – Parameters

  - Poor: "Does increasing the b parameter of BM25 increase effectiveness on corpora with longer texts?"

  - Good: "What is the effect of the b parameter of BM25 on system effectiveness on corpora with longer texts?"

  - Hypothesis: "Higher values of the b parameter of BM25 significantly increase the effectiveness on corpora with longer texts."

- ❑ Example 2 – Model Architecture

  - Poor: "Does the choice of initial retrieval model in multi-stage ranking have an effect on late stage ranking?"

  - Good: "How does the choice of initial lexical retrieval model impact the effectiveness of a late stage semantic ranking model?"

  - Hypothesis: "BM25 as initial retrieval model for a BERT based reranker performs siginifcantly worse than DirichletLM for initial retrieval."