

# SIGIR '25 notes and interesting posters

Benedikt Kantz

July 13-18 2025

## 1 Sunday

### 1.1 RAG vs. Longcontext LM Retrieval

<https://sigir2025.dei.unipd.it/detailed-program/paper?paper=f6185f0ef02dcaec414a3171cd01c697>

- LLamaIndex - framework for RAG Retrieval
- Summarization based Retrieval
  - Tree/Summarization based retrieval - construct trees and traverse by encoding tree structure
  - “RAPTOR” - flatten tree, search all nodes
- Long Context LLMs
  - Main challenge: training time  $O(n^2)$ ,  $n$  tokens
  - Extend also frequency encoding!
- focuses on LM basics with long contexts in first half

### 1.2 Efficient In-Memory Inverted Indexes: Theory and Practice

<https://sigir2025.dei.unipd.it/detailed-program/paper?paper=b38e5ff5f816ac6e4169bce9314b2996>

- Scoring/retrieval based on scores and terms
- Priming index (TODO- slides available)
- Quantization
- Reordering
- stoping (stopwords)
- Efficient Index search - dynamic pruning: WAND (weak AND) good writeup: <https://www.elastic.co/blog/faster-retrieval-of-top-hits-in-elasticsearch-with-block-max-wand>
  - Need to sort cursor when iterating (TODO: investigate further)
  - Or sort once: MaxScore, only keep relevant docs for score by precomputed upper bound for term - all combinations have combined upper bounds!
  - Can even be broken down to blocks - at the cost of storage
  - Ranker is dependent on search algorithm (e.g. BM25, LM-based, ...)
  - Threshold priming - guess initial threshold, skip docs that might not be relevant prior
- Sparse retrieval
  - Documents are naturally sparse - exploited through inverted indices, terms are related to only a few documents (except for stopwords - just remove them) – same for queries! (“Zipf’s Law”)

- Exploited through TF-IDF scoring (classics), not perfect - scoring imbalanced, lexical mismatch (semantic mismatch - e.g. F1 vs Formula One)
- ⇒ learn term weighting, sparsity, ...
- e.g. BERT for token re-weighting – learn important tokens
  - Expansion – learn new tokens to add to query, or Document Queries (i.e. Doc2Query); find similar words by using the masked prediction capabilities of BERT
  - Then: mask again, as LMs usually very dense!
  - In Practice: Query Weighting + Document Masked Expansion + Regularisation/top-k
  - But: this learned indices: distributions change, algorithms become slower – slower retrieval, sparsity reduces significantly
- ⇒ improve method by:
- \* Combine traditional BM25 retrieval for index + scoring from learned models (can even be interpolated!)
- boolean retrieval (with inverted indexes) vs. kNN/ANN (image, multimodal, semantic, ...) – future quite open!

### 1.3 Neural Lexical Search with Learned Sparse Retrieval

<https://sigir2025.dei.unipd.it/detailed-program/paper?paper=3bd4017318837e92a66298c7855f4427>  
 ⇒ seems interesting, no time (all tutorials are interesting this afternoon!) :/

### 1.4 Psychological Aspects in Retrieval and Recommendation

<https://sigir2025.dei.unipd.it/detailed-program/paper?paper=110eec23201d80e40d0c4a48954e2ff5>  
 Slides <https://github.com/aisocietylab/Psy-IR-RecSys-SIGIR25/>

- IR systems integrated in life - play role in life - psychological impact (cognition, decision, ...)!  
 ⇒ model human using emotions/personality/overload/mood (term: psychological states!)
- “Cognitive architecture”: first IR paper in cognitive science journal (by CS researchers!) - go back to roots
- higher level/lower level processes (e.g. attention)
- Key processes
  - Attention: pull focus on key elements
  - Memory: memorize focused elements ⇒ pull focus
  - Decision-making: (ir-)rationality, heuristic-based
- Map processes between IR/RS scenarios

IR/RS Task	Cognitive Process
Query (Re-)formulation	Working memory, attention
Ranking, Interface Design	Attention, perception
Session continuation or abandonment	Episodic memory, fatigue
Item selection	Decision-making, heuristics
Personalization (over time)	Long-term memory, learning

- Cognitive architectures: are “computation frameworks” – helps us to apply models & test them, align models in IR with users!
- Symbolic architecture
  - Discrete symbols, with rules operating on knowledge

- ⇒ explainable, explicitly encoded; hard to expand due to handcrafted rules
- Emergent/Connectionist architectures
  - across network/graph, parallel processing, nodes represent subsymbolic elements
  - ⇒ not (or limited) explainability!, a lot of data needed
- Hybrid: more flexibility
- Survey: <https://link.springer.com/article/10.1007/s10462-018-9646-y#Fig3>
- ACT-R: Adaptive Control of Thought – Rational (ACT-R): memory types (a la von Neumann), with rules inside
- Application: recency bias in retrieval, e.g. music retrieval for non-mainstream music
- SOAR: goal oriented, episodic memory – agentic framework?
- CLARIOn: hybrid
- LIDA: Global Workspace Theory, consciousness vs unconscious focus
- Problems: difficult to integrate in fuzzy pipelines like DL models, quite interdisciplinary ...
- Cognitive Biases
  - biases play against rational (logical and optimal decisions)
  - influences in algorithmic decision making; encoded in various tasks, LM models (generative AND embedding!)
  - Feature-Positive Effect: not aware of *missing* things: which information is ignored; e.g. which aspects of a job RecSys are ignored?
  - Social/Cultural Homophily: users with similar traits like similar things: experiments show users like music from their own country.
  - Conformity bias: conform to opinion of others; e.g. show users ratings of other users first biases – could be used to lead towards higher/lower ratings?
  - Declinism: belief, world was a better place – found in increase in negative terms in music lyrics
  - Primacy/Recency Bias
- Personality and affect
  - States & Traits
  - “Affective Computing”: detect emotions from metadata/interactions

## 2 Monday

### 2.1 BM25 and All That

- History of IR, back to Cranfield experiments, first probabilistic theories and how they theorized/implemented BM1-& BM25

### 2.2 Multimodal Retrieval

#### 2.2.1 DePro: Domain Ensemble using Decoupled Prompts for Universal Cross-Domain Retrieval

- Prompt tuning/embedding optimisation for concatenated prompts

### 2.3 Reproducibility in Domain-Specific, Multimodal, ... Retrieval

(the critic's section)

### 2.3.1 Benchmark Granularity

- Image-Text: bidirectional problem
- Problem: benchmarks very curated/clean, real-world: noisy, ...
- Eval: shuffle words, different granularities of descriptions
- Results:
  - performance increases if queries are more fine-grained  $\implies$  miss out on performance of model, use multiple granularities in benchmarks!
  - perturbations usually hurt
- Very inconsistent across models/datasets!

### 2.3.2 Visual Document Retrieval + Late Interactions

- ViT: patch-based encoding of document (ColPali/ColQwen2 <https://huggingface.co/vidore/colqwen2-v1.0>) using ColBERT paradigm (offline search using MaxSim <https://github.com/stanford-futuredata/ColBERT>)
- Image-Documents has less loss in performance even in big index sizes compared to OCR
- Reason: number of patches/text tokens factor is relevant
- Matching to visual tokens to query is high – seems effective

### 2.3.3 Assessing effective token length of multimodal retrieval

- Information Alignment: CLIP has a modality gap between image and text!
  - Comes from short/descriptive captions
  - Long-Text performance/beyond token budget using pooling?
  - Domain-Specific Datasets
- $\implies$  no model could use full token budget!, also variance between datasets
- How was the token length varied? removing adjectives, just cut? – was just cut... (maybe information afterwards not relevant/good?)

### 2.3.4 Refined Medical Search via Dense Retrieval and User Interaction

Interesting for A+CHIS/HEREDITARY search engine?

- Enhance User Experience with Interaction integration in Dense (embedding) Retrieval
- TripClick IR Dataset - $\downarrow$  A+CHIS?
- Fine tune PubMedBERT on TripClick Triplets (Document, D+, D-) – what are D+/D-
- Retrieval by embedding example queries from TripClick dataset + adding similar documents/queries from dataset of session to query

### 2.3.5 Reassessing LLM for Boolean Queries for Systematic Reviews

- In medical domains: a lot of very complex boolean queries for literature reviews
- Prior paper: boolean queries for systematic reviews, other reproducibility paper found: their results do not generalize – but did not use syntax checker, produces wrong results!
- Also: did not use guide-documents for prompt engineering
- Syntax check - regenerate
- Q: Did they use grammar for queries (constrained generation?) json output can reduce performance of LLMs <https://arxiv.org/abs/2408.02442>

### 2.3.6 Revisiting Algorithmic Audits of TikTok: Poor Reproducibility and Short-term Validity of Findings

- TikTok audits are difficult to reproduce (change of platform/bot banning)
  - Personalization stronger for longer videos
  - Country ID by Proxies
- ⇒ audits only short-term valid

### 2.3.7 Beyond Reproducibility: Advancing Zero-shot LLM Reranking Efficiency with Set-wise Insertion

- Previous: Comparison by “Bubble-Sort-by-LLM”
- Do chunked reordering using LMs

## 2.4 Knowledge & KGs

### 2.4.1 Segmentation Similarity Enhanced Semantic Related Entity Fusion for Multi-modal Knowledge Graph Completion

- Using ViT+Semantic Segmentation to complete KGs with more multimodal knowledge (location based on images, ...)
- Transformers over Transformers...

### 2.4.2 Mitigating Modality Bias in Multi-modal Entity Alignment from a Causal Perspective

- finding equality/alignment between KGs using semantic information

## 2.5 Human and Interfaces

### 2.5.1 The In-Situ Effect of Offensive Ads on Search Engine Users

- people see avg. 5000 ads/day
- offensive ads more effective, but at a risk – evoke negative emotions
- complaints: usually only very few offensive ads (Bing: a single offensive ad spreadsheet!)
- then: user study to count (offensive) ads
- Result: users can barely distinguish ads, irrelevance and offensiveness is very correlated; irrelevance is considered more offensive than offensive itself!
- People do not recall offensive ads on bing/google

## **2.5.2 Simulating Before Planning: Constructing Intrinsic User World Model for User-Tailored Dialogue Policy Planning**

- Agentic framework ...

# **3 Tuesday**

## **3.1 Keynote: Digital Health**

- Vaguely related to IR, advertisement for the industrial lab
- Prediction of treatment/preselection of people based on disease history
- Kenneth Church: Chatbots are not the solution for Callcenters - could be just an app...

## **3.2 Search and Ranking**

(Peter's favorite session!)

### **3.2.1 CoDIME: A Counterfactual Approach for Dimension Importance Estimation through Click Logs**

- Dimension Importance Model Estimation (DIME): only keep good dimensions for dense retrieval
- Estimate through LLM Gen docs (inconsistent) OR relevant docs from collection (need ground truth)
- Solution: use implicit feedback (click data)
- Estimate Correlation between rankings/clicks; fit linear model between this correlation (for dimension) as estimator
- Improves retrieval significantly - interesting for dense semantic retrieval!

### **3.2.2 Stitching Inner Product and Euclidean Metrics for Topology-aware Maximum Inner Product Search**

~~StitchingInnerProductandEuclideanMetricsforTopology-awareMaximumInnerProductSearch~~

- Nearest neighbour search using euclidian distance is not great, as inner poroduct not a natural norm (no triangle ineq); need it for speedup of MIPS
- This seems like a convex hull over neighbourhood?

### **3.2.3 On the Scaling of Robustness and Effectiveness in Dense Retrieval**

- Robustness of models largely ignored, lower limit of system relevant (in OOD settings)!
- Incorporate Robustness into loss
- Danger zone: did they use the test data to compute the robustness in the loss? – not a great answer in session
- Also no comparison to BM25 baseline...

### 3.2.4 Optimizing Compound Retrieval Systems

- Cascading retrieval systems (rerankers...)
- LLM Ranking does not really fit into reranker approach
- New paradigm: compound retrieval: *any combination* – cascading just a specific instance
- Propose fully differentiable rerank pipeline - can just learn to design the compound system
- Results: dynamic trade-off between LM calls and
  - BM25 ranker for pointwise comparison
  - Cascading system seems to be good choice
  - sparse selection of top seems also nice (or even other patterns of selection)
- Conclusion: could be interesting for novel compound systems

### 3.2.5 Hypencoder: Hypernetworks for Information Retrieval

- Problem: the linearity of dot product could result in non-linearly-sperarable subspaces for retrieval
- Solution: non-linear query-specific NN using Hypernetworks - learned NN that produces weights for other NNs

### 3.2.6 Classifying Term Variants in Query Formulation

- Variety of query for the same goal – 95/100 queries unique even for simple query!
- over 10k query formulations for 100 query back stories
- Analysis: remove back story; analyse added terms – users added 70% terms by themselves
- User modify semantics, misspell...
- Or enrich terms with information types or sources

### 3.2.7 Locality-Sensitive Indexing for Graph-Based Approximate Nearest Neighbor Search

- kNN search: graph based, often fsated by searching topologies
- Update patterns: reconnection best, deletion/hiding hurt performance (recall/time)
- LIGS: bucketize/hash vectors together – simulated graph by overlaying buckets – connections only implicit!
- good for high-churn applications (social media/networks)

### 3.2.8 Constrained Auto-Regressive Decoding Constrains Generative Retrieval

- Generative retrieval methods use constrained generative retrieval
- Either: step-wise elimination or beam-search over  $k$  sampling steps
- Result: erroneous retrieval, and will give high precision and low recall (just as we observed in the GutBrainIE evaluation!!)

## 3.3 Biomedical and Health

### 3.3.1 Cooking with Conversation: Enhancing User Engagement and Learning with a Knowledge-Enhancing Assistant

- Allow users either an active or passive guidance for Cooking
- include annotated recipes (KG/facts)

### 3.3.2 General Neural Embedding for Sequence Distance Approximation

- Usually: Distance metrics like edit distance hard to compute
- distance approximated using NN, existing approaches lack generality
- Their Approach: Small domain head, shared CNN+Transformer block for all domains!
- Show good performance on both DNA and GPS data (??), no significance test

### 3.3.3 ProtChatGPT: Towards Understanding Proteins with Hybrid Representation and Large Language Models

- PLP-Former for special protein tokens

### 3.3.4 Combining Evidence and Reasoning for Biomedical Fact-Checking

- 3-stage framework for verification of facts, including justification aoutput of LMs

## 4 Wednesday

### 4.1 Keynote: Please meet AI, our dear new colleague. In other words: can scientists and machines truly cooperate?

(Iryna Gurevych)

- Peer review assistance through various tools: ACL and AAAI already incorporate this!
- Specific Review Datasets
- Workflow for Review Process, AI cannot fully arrive at review but help prefill fields
- Related works: difficult to write a good argumentative section, not just summarize!
- Connection between reviews and revisions – hard, as the task is quite undefined!
- Aspects in review: Detect aspects in fields semi-autonomously (ontology!); LLM generated reviews are very similar w.r.t. aspects! – use for detection of LLM detection;
- Lazy Review heuristic: i.e. did not do ...
- Reviews and promised changes: relate them automatically, verify completion!

### 4.2 Efficiency

#### 4.2.1 WebANNS: Fast and Efficient Approximate Nearest Neighbor Search in Web Browsers

(Peter: nice WASM usage); very good presentation!

- Previously: USed IndexedDB within browser / prefetch data points, us HNSW + JS (?)
- Proposed: WASM, problem WASM  $\rightleftharpoons$  IndexedDB, IndexedDB slow
- Tree In-Memory, rest in IndexedDB+JS (three-tier data storage: WASM-JS-DB), Lazyload per tree level
- Benchmarks show sub-second performance!

#### 4.2.2 TITE: Token-Independent Text Encoder for Information Retrieval

- Reduce computational cost by aggregating tree-wise, but the model can increase dimensionality
- Related to Funnel Transformers: pool (e.g. 2 $\times$ ) over one of the attention dimensions
- without loss of performance!

#### **4.2.3 WARP: An Efficient Engine for Multi-Vector Retrieval**

(Best Paper Award)

- Combine XTR+ColBERTv2/PLAID to do token-wise retrieval

#### **4.2.4 An Analysis on Matching Mechanisms and Token Pruning for Late-interaction Models**

- Prune tokens based on attention/...
- Little efficiency loss

#### **4.2.5 Data Augmentation for Sample Efficient and Robust Document Ranking**

- Augment data by splitting docs & rerank passages of docs based on simple methods
- Solid eval on various loss/augmentation combinations
- Augmentation is useful, especially in low-resource
- Acts as regularizer, allows transfer from e.g. MSMARCO to other domains

### **4.3 Reranking**

#### **4.3.1 Bridging Personalization and Control in Scientific Personalized Search**

- Focus: Reranker with personalized search
- Cross-Encoder for reranking, their contribution: personalization
- Approach: represent history as context to the reranker; add personalization by mixing concepts using 'routing'
- Could be interesting to visualize within OnSET to improve user guidance?

#### **4.3.2 Reason-to-Rank: Distilling Direct and Comparative Reasoning from Large Language Models for Document Reranking**

- Add reasoning model for reranking model

### **4.4 Low-Resouce**

#### **4.4.1 When Less is Enough: Optimizations for Low-Cost Recommendation Systems**

- Low-ARPU regions: profitability @ \$4/user

#### **4.4.2 Efficient Approximate Nearest Neighbor Search on a Raspberry Pi**

#### **4.4.3 IR for AAC Users: A Hyperdimensional Computing (Vector Symbolic Architectures) Approach**

#### **4.4.4 Some Things Never Change: Overcoming Persistent Challenges in Children IR**

## **5 Thursday (Workshops)**

### **5.1 ReNeuIR at SIGIR 2025: The Fourth Workshop on Reaching Efficiency in Neural Information Retrieval**

## **6 Posters**

Table 1:

Poster	Information																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
<p><b>Resource for Error Analysis in Text Simplification: New Taxonomy and Test Collection</b></p> <p>Benjamin Venneker[1,2,3], Diana Ernakeva[2], Pierre De Loor[3,4]</p> <p><b>Introduction</b></p> <p>The analysis of text simplification errors is a key step in understanding the error made by our system. Many activities have been made in this field, including error detection and mitigation systems, or simple test collections. The main bottleneck of LSTF progress for multiple years: tasks differed greatly in quality, difficulty, and complexity of texts. However, we observed that many generated publications are irrelevant because they do not reflect the real needs of end-users. In this paper, we propose a new direct method for ATs tasks progress, new resources, and new taxonomy.</p> <p><b>Taxonomy</b></p> <p>We propose a new taxonomy for ATs errors based on their nature. We distinguish between three main categories: <b>Textual</b>, <b>Lexical</b>, and <b>Syntactic</b>. Each category is further divided into sub-categories. For example, the <b>Textual</b> category includes <b>Text length</b>, <b>Text complexity</b>, and <b>Text readability</b>.</p> <p><b>Dataset</b></p> <table border="1"> <thead> <tr> <th>Category</th> <th>Sub-category</th> <th>Count</th> <th>Mean</th> <th>SD</th> </tr> </thead> <tbody> <tr> <td rowspan="10">Textual</td> <td>Text length</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text complexity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text readability</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text coherence</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text fluency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text consistency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text clarity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text precision</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text accuracy</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Text specificity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td rowspan="10">Lexical</td> <td>Vocabulary size</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word frequency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word length</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word complexity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word readability</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word coherence</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word fluency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word consistency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word clarity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Word precision</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td rowspan="10">Syntactic</td> <td>Syntax complexity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax readability</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax coherence</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax fluency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax consistency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax clarity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax precision</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax specificity</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax accuracy</td> <td>2000</td> <td>100</td> <td>10</td> </tr> <tr> <td>Syntax consistency</td> <td>2000</td> <td>100</td> <td>10</td> </tr> </tbody> </table> <p><b>Simpletext 2020 results</b></p> <table border="1"> <thead> <tr> <th>Model</th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> <th>K</th> <th>L</th> <th>M</th> <th>N</th> <th>O</th> <th>P</th> <th>Q</th> <th>R</th> <th>S</th> <th>T</th> <th>U</th> <th>V</th> <th>W</th> <th>X</th> <th>Y</th> <th>Z</th> </tr> </thead> <tbody> <tr> <td>DebtCollector</td> <td>0.78</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.73</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.84</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.85</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.86</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.87</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.88</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.89</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.90</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.91</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.92</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.93</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.94</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.95</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.96</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.97</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.98</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>0.99</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> <tr> <td>DebtRecovery</td> <td>1.00</td> <td>0.98</td> <td>0.74</td> <td>0.19</td> <td>0.03</td> <td>0.30</td> </tr> </tbody> </table> <p><b>Conclusion</b></p> <p>We introduced and evaluated the first taxonomy of errors in text simplification. Our taxonomy and the corresponding test collection can be used to evaluate and compare different models. Our results show that the proposed taxonomy is effective in identifying different types of errors and that it can be used to improve the quality of text simplification systems.</p> <p><b>Future work</b></p> <p>In the future, we plan to extend our taxonomy to cover more types of errors and to include more datasets. We also plan to evaluate our taxonomy on different datasets and to compare its performance with other taxonomies.</p>	Category	Sub-category	Count	Mean	SD	Textual	Text length	2000	100	10	Text complexity	2000	100	10	Text readability	2000	100	10	Text coherence	2000	100	10	Text fluency	2000	100	10	Text consistency	2000	100	10	Text clarity	2000	100	10	Text precision	2000	100	10	Text accuracy	2000	100	10	Text specificity	2000	100	10	Lexical	Vocabulary size	2000	100	10	Word frequency	2000	100	10	Word length	2000	100	10	Word complexity	2000	100	10	Word readability	2000	100	10	Word coherence	2000	100	10	Word fluency	2000	100	10	Word consistency	2000	100	10	Word clarity	2000	100	10	Word precision	2000	100	10	Syntactic	Syntax complexity	2000	100	10	Syntax readability	2000	100	10	Syntax coherence	2000	100	10	Syntax fluency	2000	100	10	Syntax consistency	2000	100	10	Syntax clarity	2000	100	10	Syntax precision	2000	100	10	Syntax specificity	2000	100	10	Syntax accuracy	2000	100	10	Syntax consistency	2000	100	10	Model	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	DebtCollector	0.78	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.73	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.84	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.85	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.86	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.87	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.88	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.89	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.90	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.91	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.92	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.93	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.94	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.95	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.96	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.97	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.98	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	0.99	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	DebtRecovery	1.00	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
Category	Sub-category	Count	Mean	SD																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Textual	Text length	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text complexity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text readability	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text coherence	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text fluency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text consistency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text clarity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text precision	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text accuracy	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Text specificity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Lexical	Vocabulary size	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word frequency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word length	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word complexity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word readability	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word coherence	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word fluency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word consistency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word clarity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Word precision	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Syntactic	Syntax complexity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax readability	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax coherence	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax fluency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax consistency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax clarity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax precision	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax specificity	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax accuracy	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	Syntax consistency	2000	100	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Model	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
DebtCollector	0.78	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.73	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.84	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.85	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.86	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.87	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.88	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.89	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.90	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.91	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.92	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.93	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.94	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.95	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.96	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.97	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.98	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	0.99	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DebtRecovery	1.00	0.98	0.74	0.19	0.03	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				

Continued on next page

Table 1: (Continued)

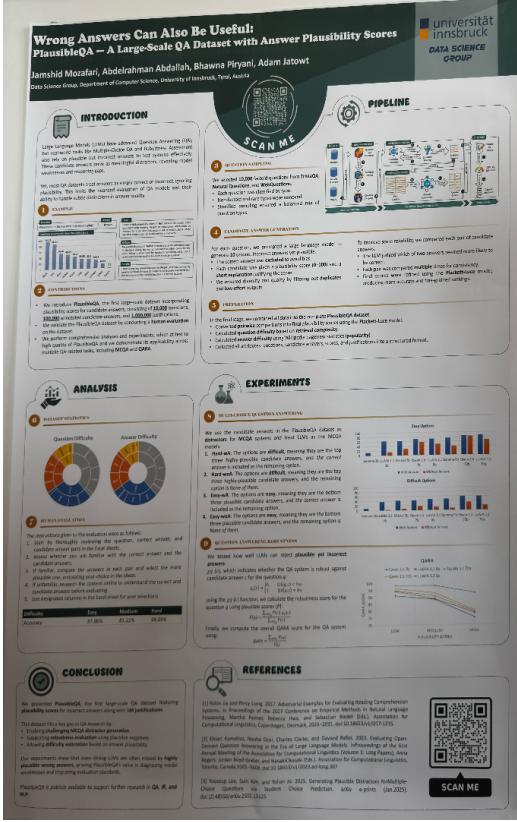


**ROKSANA: An Open-Source Toolkit for Robust Graph-Based Keyword Search**

The ROKSANA toolkit is an open-source solution for robust graph-based keyword search. It includes a Data Management Toolkit, a Search Engine Toolkit, and an Online Playground. The Data Management Toolkit provides tools for managing datasets, including a Data Pipeline, Data Cleaning, and Data Transformation. The Search Engine Toolkit includes a Search Engine API, a Search Engine UI, and a Performance Analysis tool. The Online Playground allows users to experiment with the toolkit's features.

**ROKSANA: An Open-Source Toolkit for Robust Graph-Based Keyword Search *equeries, gold set, resus before, k. values-15, 10, 201***

<https://doi.org/10.1145/3726302.3730154>



**Wrong Answers Can Also Be Useful: PlausibleQA — A Large-Scale QA Dataset with Answer Plausibility Scores**

PlausibleQA is a large-scale dataset for question answering (QA) that includes answer plausibility scores. The dataset contains over 100,000 annotated candidate answers and 1,000,000 justifications. The interface includes sections for INTRODUCTION, ANALYSIS, EXPERIMENTS, and CONCLUSION. The ANALYSIS section shows donut charts for question and answer difficulty. The EXPERIMENTS section includes a pipeline diagram and a bar chart comparing QNLP and QG models.

**Wrong Answers Can Also Be Useful: PlausibleQA — A Large-Scale QA Dataset with Answer Plausibility Scores 100,000 annotated candidate answers, and 1,000,000 justifications.**

<https://doi.org/10.1145/3726302.3730299>

Continued on next page

Table 1: (Continued)

**Measuring the Fairness Gap between Retrieval and Generation**  
 Sandeep Avula, Chia-Jung Lee, Rongting Zhang, Vanessa Murdock  
 AWS AI/ML Responsible AI

• Fairness degrades consistently from retrieval to generation  
 • Create and *understand* tasks with larger contexts produce more fair results than *evaluate* tasks  
 • Larger models and larger context sizes produce more fair results

**Conversational Search: Towards Personalization and Evaluation** the generated text  $G = (5, \dots, Sm)$  with attribution weights  $c_i$ ; and  $E$ ,  
<https://doi.org/10.1145/3726302.3730126>

**101 REANIMATOR**  
**REANIMATE RETRIEVAL TEST COLLECTIONS WITH EXTRACTED AND SYNTHETIC RESOURCES**  
 Björn Engelmann, Fabian Haak, Philipp Scher, Mani Erfanian Abdoust, Linus Netze, Meik Bikowski

**Limitations of Legacy Test Collections**

- Limited scope: often static & un-modifiable
- Generalizability gap: task-specific & narrow
- Extraction bias: full texts, tables, figures
- Limited scope: often static & un-modifiable
- Generalizability gap: task-specific & narrow
- Extraction bias: full texts, tables, figures

**Our Solution: REANIMATOR**  
 A versatile framework to reanimate existing test collections to serve as training with extracted and synthetic resources

**The REANIMATOR Pipeline**

**Use Case: Tables in RAG**

**Results and Evaluation**

**Example: TREC-COVID+**

- Input Corpus: TREC-COVID+ COVID-19 test collection, involving 100k in TREC format.
- Extracted: full texts, tables, figures
- Synthetic resource assessment of paragraphs and tables
- Validation: human evaluation on 100k documents
- Synthetic resource assessment of paragraphs and tables
- Validation: human evaluation on 100k documents
- Synthetic resource assessment of paragraphs and tables
- Validation: human evaluation on 100k documents
- Requirements: runs on regular computer

**REANIMATOR is available now!**

**CIR** **SMC** **Technology Arts Sciences TH Köln**

**REANIMATOR: Reanimate Retrieval Test Collections with Extracted and Synthetic Resources** Björn Engelmann, Fabian Haak, Philipp Scher, Mani Erfanian Abdoust, Linus Netze, Meik Bikowski

<https://doi.org/10.1145/3726302.3730342>

Continued on next page

Table 1: (Continued)

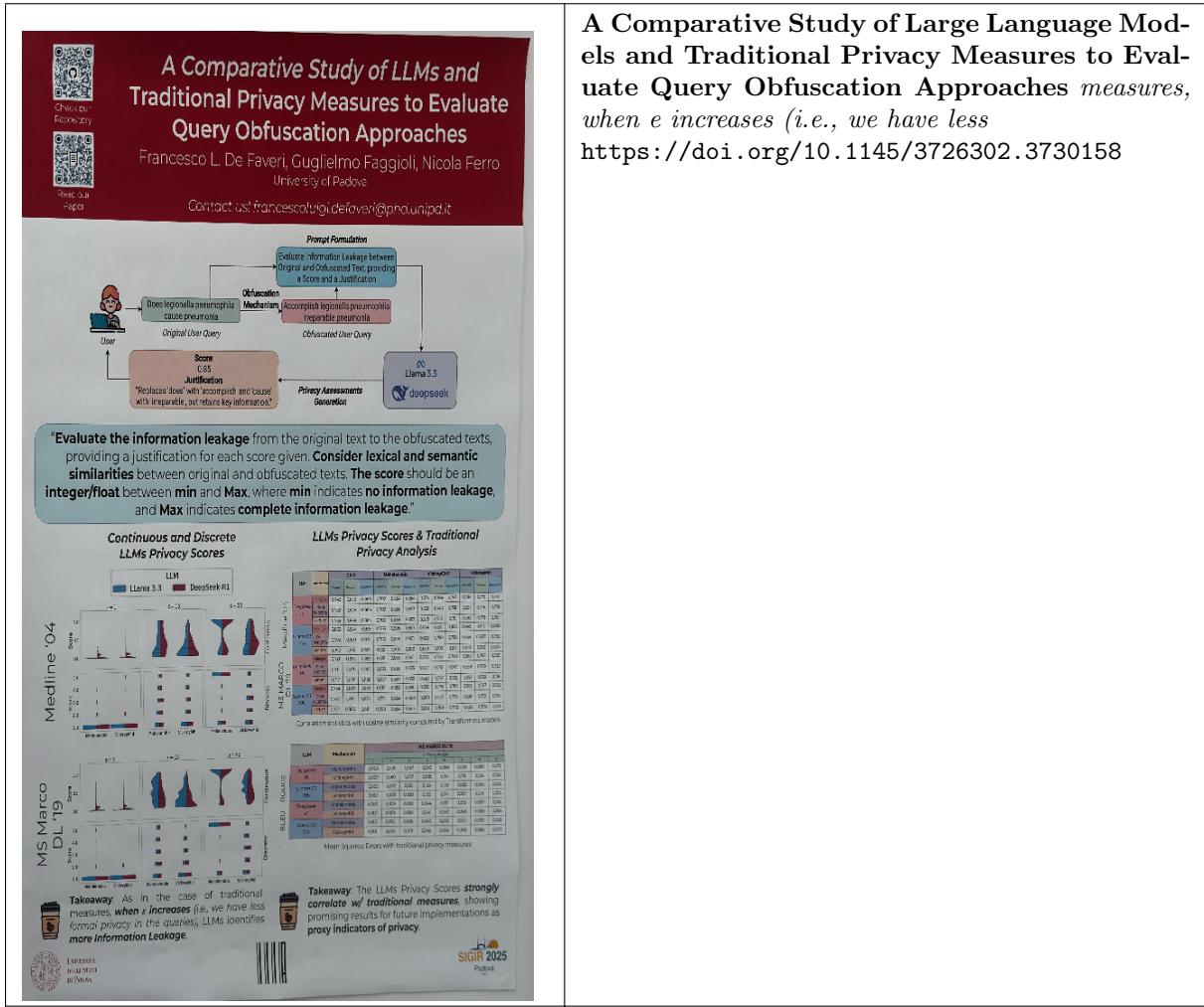
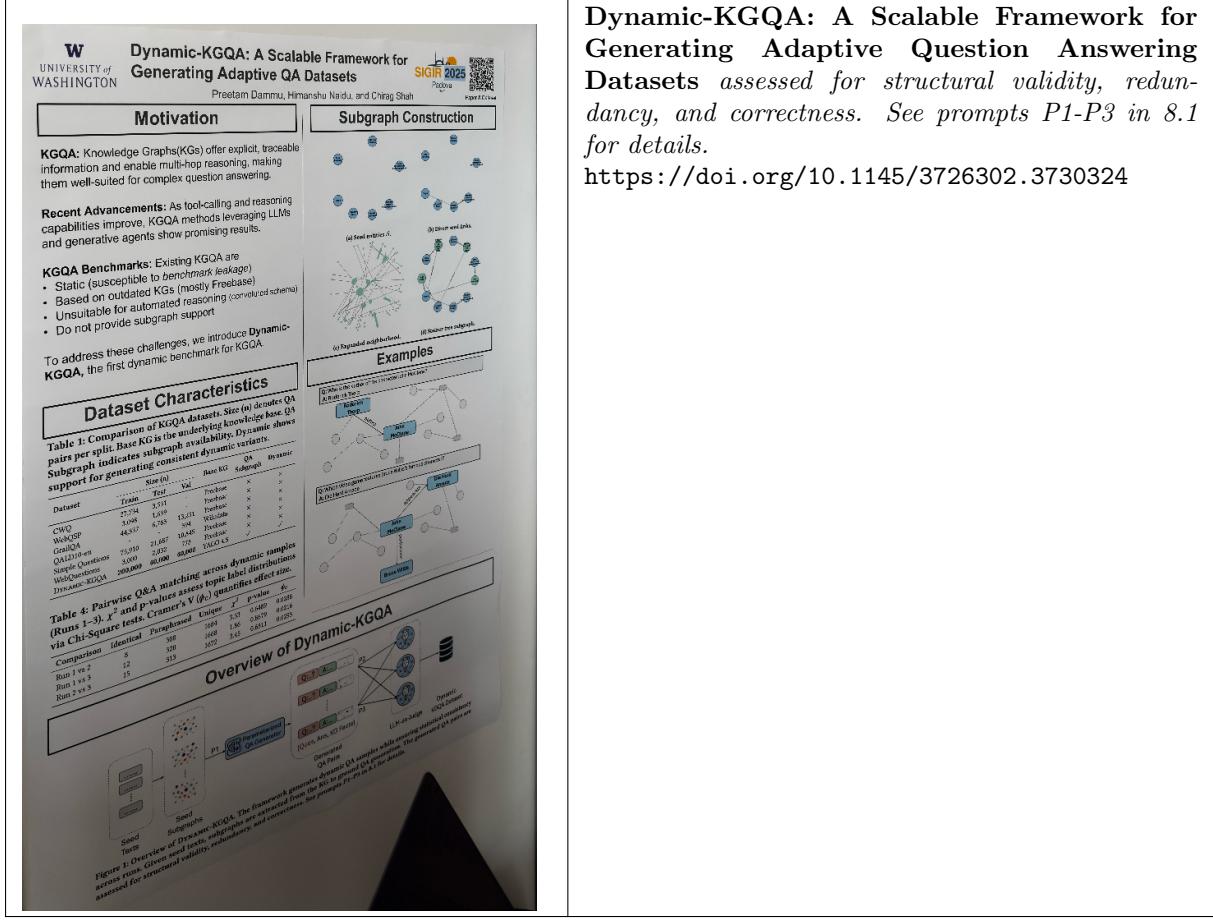


Table 1: (Continued)



Continued on next page

Table 1: (Continued)

**U-Sticker: A Large-Scale Multi-Domain User Sticker Dataset for Retrieval and Personalization**

Heng Li<sup>1,2</sup>, Meitali Chee<sup>1,2</sup>, Jiany Wang<sup>1</sup>, Weizhi Ma<sup>1</sup>, Zhiqiang Gao<sup>1</sup>, Mai Zhang<sup>1,2\*</sup>, Quan Cheng<sup>1,2</sup>

<sup>1</sup>DCST, Tsinghua University, Beijing, 100084, China <sup>2</sup>Quan Cheng Laboratory<sup>2</sup> AIR, Tsinghua University, Beijing, 100084, China

**Introduction**

**Dataset Construction**

**Application 1: User Behavior Analysis**

- Users have distinct styles.
- Users display cross-domain behavior change.

**Application 2: Sticker Recommendation**

**Fig. 1: User Behavior Analysis**

**Fig. 2: Sticker Features for 10 Domains**

**Fig. 3: Some users' sticker preferences change over time (User C, from 10 to 8 years)**

**Fig. 4: Users' behavior differ to different people**

**Fig. 5: Stable performance with room for improvement. Domain-wise demonstrate contrast. Each domain is independent, static and weightless**

### The Agent Perspective In LLM-Based Strategic Information Retrieval Ecosystems 'DCST, Tsinghua University, Beijing, 100084, China 2 Quan Cheng Laboratory2 AIR, Tsinghua University, Beijing, 100084, China

<https://doi.org/10.1145/3726302.3730125>

Continued on next page

Table 1: (Continued)

**TINK: TEXT INFORMATION NAVIGATION KIT 3. X,  $= [X_0] vev$  is a set of texts such that  $VX, E Xv, X, \in Z$**

<https://doi.org/10.1145/3726302.3730141>

**Tip of the Tongue Query Elicitation for Simulated Evaluation rank correlations gave moderately high values (e.g.,  $r = 0.9142$  in Landmark,**

<https://doi.org/10.1145/3726302.3730335>

**Motivation & Research Questions**

- Tip of the Tongue (TOT) search system aims to assist users in solving TOT queries while it expresses user’s intent in a more accurate, less formal, and flexible manner. This work covers over 2 TOT search engines: CQA and QAC.
- Heavy reliance on community question answering (CQA) websites.
- Domain Specific e.g., movies, books.
- No Stability: latency induced connection drops, unreliable connections, etc.
- Challenge in coverage → user query in development.
- Solution: Simulated Evaluation (via LLMs + Human).
- Research Questions:

  - RQ 1: Can TOT queries from LLMs be effective and evaluate of LLM’s retrieval system?
  - RQ 2: Can we elicit TOT queries from humans to measure query coverage?
  - RQ 3: Can the elicited TOT queries methods work in other domains underrepresented in CQA-collected test collections?

**To Validate Elicited TOT Queries?**

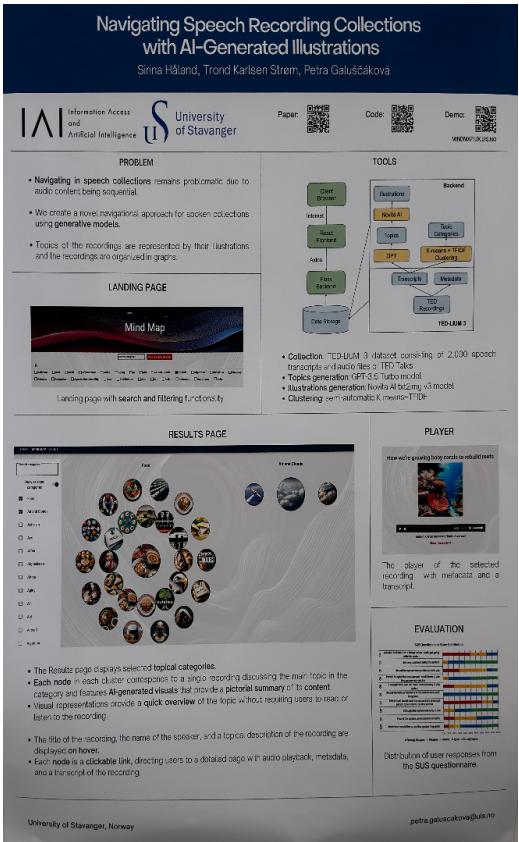
Figure 2 shows the TOT query elicitation process using CQA datasets and related tasks. It includes ground truth performance (MRR@10 and NDCG@10) and Pearson correlation to assess the agreement between the elicited TOT queries.

**Takeaways & Outcomes**

- 100 million synthetic local TOT-based datasets by introducing LLM and human-related queries. Over 94% available and retrievable in TOT dataset.
- Our effort removes the need for manual testing, avoids biasing evaluations, and mitigate domain bias.
- PLCC@1000: Retrieved TOT queries in under 1500 per domain.
- Code Resource: Open-source code for LLM evaluation and the human TOT query elicitation.

Continued on next page

Table 1: (Continued)

 <p><b>Navigating Speech Recording Collections with AI-Generated Illustrations</b> Sirina Häland, Trond Karlsen Strom, Petra Galuscáková</p> <p><b>PROBLEM</b></p> <ul style="list-style-type: none"> <li>• Navigating in speech collections remains problematic due to audio content being sequential.</li> <li>• We create a novel navigation approach for spoken collections using generative models.</li> <li>• Topics of the recordings are represented by their illustrations and the recordings are organized in graphs.</li> </ul> <p><b>LANDING PAGE</b></p> <p>Landing page with search and filtering functionality.</p> <p><b>TOOLS</b></p>  <p><b>RESULTS PAGE</b></p> <p>The Results page displays selected topical categories.</p> <ul style="list-style-type: none"> <li>• Each node in each cluster corresponds to a single recording discussing the main topic in the category and features AI-generated visuals that provide a pictorial summary of its content.</li> <li>• Visual representations provide a quick overview of the topic without requiring users to read or listen to the recording.</li> <li>• The title of the recording, the name of the speaker, and a topics description of the recording are displayed on hover.</li> <li>• Each node is a clickable link, directing users to a detailed page with audio playback, metadata, and a transcript of the recording.</li> </ul> <p><b>PLAYER</b></p> <p>The player of the selected recording with metadata and a transcript.</p> <p><b>EVALUATION</b></p>  <p>Distribution of user responses from the SUS questionnaire.</p>	<p><b>Navigating Speech Recording Collections with AI-Generated Illustrations</b> <i>Sirina Häland, Trond Karlsen Strom, Petra Galuscáková</i></p> <p><a href="https://doi.org/10.1145/3726302.3730136">https://doi.org/10.1145/3726302.3730136</a></p>
---	---