Unified Summary Table of Academic Papers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | Title | Authors | Methods Used | Dataset / Domain | Performance Measure | Input / Task |
| 1 | Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks | Patrick Lewis et al. | DPR + BART (RAG-Sequence & RAG-Token) | Wikipedia, NQ, TriviaQA | ** EM: 44.5%, F1: 58.1%, BLEU, ROUGE QA-based text input** | QA-based text input |
| 2 | RAG for Academic Literature Navigation | Ahmet Yasin Aytar et al. | 5-stage RAG pipeline (GROBID, Abstract-first retrieval) | arXiv papers, textbooks | ** RAGAS - Context: 78%, Faithfulness: 72%, Relevance: 75% Academic queries** | Academic queries |
| 3 | Benchmark of PDF Information Extraction Tools | Norman Meuschke et al. | Levenshtein-based benchmarking (GROBID, CERMINE) | DocBank (500K arXiv pages) | ** Precision: 93%, Recall: 91%, F1: 92.5%, Accuracy: 94% PDF academic docs** | PDF academic docs |
| 4 | S2ORC: Open Research Corpus | Kyle Lo et al. | Metadata extraction + BERT eval | 81.1M papers (PubMed, arXiv) | ** Accuracy: 96.1%, F1: 95.4% Raw PDFs, LaTeX** | Raw PDFs, LaTeX |
| 5 | Survey on AI-Generated Plagiarism | Shushanta Pudasaini et al. | Literature survey + experiment (ChatGPT, QuillBot) | ArguGPT, HC3, CHEAT | ** Accuracy: 85%, Precision: 82%, Recall: 80%, AUROC: 88% Human vs AI-generated text** | Human vs AI-generated text |
| 6 | Ontology + DL for Legal Texts | Yong Ren et al. | Ontology modeling + Deep Learning | Chinese legal documents | ** Accuracy: ~90%, Qualitative Legal PDFs** | Legal PDFs |
| 7 | AI in Higher Education Research | I. Cingillioglu et al. | Automated RCT system | 1193 online learners | ** Engagement ↑: 20%, Automation: qualitative Web-based learning data** | Web-based learning data |
| 8 | RAG-Based LLM Systems from PDFs | Ayman A. Khan et al. | Chunking, Embedding, RAG over PDFs | Real-world business/academic PDFs | ** User Feedback: 88% satisfaction Uploaded PDFs** | Uploaded PDFs |
| 9 | Extract-and-Abstract Summarization | Yuping Wu et al. | Shared encoder-decoder, saliency masking | CNN/DailyMail | ** ROUGE-1: 43.8%, ROUGE-2: 20.7%, ROUGE-L: 40.2% Summarization input** | Summarization input |
| 10 | State-of-the-Art in Academic Plagiarism | Norman Meuschke, Bela Gipp | Survey (character, citation, cross-lingual) | 376 papers (1994–2019) | ** Simple Detection: 88–96%, Paraphrased: <70% Academic essays** | Academic essays |
| 11 | Plagiarism in AI-Generated Research | R. Krishna et al. | Semantic review + expert judgment | Manually created AI papers | ** Expert Originality: 60–75% AI-generated academic content** | AI-generated academic content |
| 12 | Software Plagiarism in Age of AI | T. Sağlam, L. Schmid | GPT-4 code obfuscation test | AI-generated code (GitHub, Copilot) | ** Detection Accuracy: 85% (human), 50–65% (AI) Code samples** | Code samples |
| 13 | Semantic Analysis for Relation Extraction | M.-T. C. Evans et al. | Transformer + semantic features | TACRED, NLP datasets | ** F1: 74.2%, Precision: 76.1%, Recall: 72.3% Text pairs** | Text pairs |
| 14 | Efficient Vector Similarity Search | K.-U. Sattler et al. | Cosine similarity, NN search | Simulated doc collections | ** Accuracy: 96%, Speed: ~30ms/query Vectorized segments** | Vectorized segments |
| 15 | PlagBench: Duality of LLMs | J. Lee et al. | LLM generation + detection (PlagBench) | Benchmark of real vs AI text | ** Precision: 87%, Recall: 83%, F1: 85% AI vs human text** | AI vs human text |
| 16 | TAISR Framework for IS Research | Benjamin Ampel et al. | TAISR + case studies | IS research docs | ** Utility Score: 92%, Qualitative analysis Advanced LLM tasks** | Advanced LLM tasks |
| 17 | Prompt Pattern Catalog | Dominic Widdows et al. | 13 reusable prompt patterns | GPT-4, Claude, Gemini | ** Prompt Reliability: >80% (qualitative) Prompt templates in SDLC** | Prompt templates in SDLC |
| 18 | Chain-of-Thought Prompting | Jason Wei et al. | CoT prompting (few-shot with reasoning chains) | GSM8K, SVAMP, CSQA, etc. | ** Accuracy on GSM8K: 75% Reasoning tasks** | Reasoning tasks |
| 19 | Instruction Fine-tuning + Merging | Yizhong Wang et al. | Model merging (Llama3, Phi-3, Mistral) | Financial text classification | ** Zero-shot gain: +12%, Task-specific: +19% Financial NLP tasks** | Financial NLP tasks |
| 20 | Residual Memory Transformer | Hanqing Zhang et al. | Plug-in for control over CLMs | GPT-2, CTG benchmarks | ** Efficiency: +30%, Controllability ↑ Controlled text generation** | Controlled text generation |

Table Performance Analysis Table: Academic Research Assistant Using RAG

We compiled a performance analysis table of 20 academic papers that are highly relevant to our Academic Research Assistant project. Each paper was examined for its methods, datasets, performance metrics, and task focus—such as question answering, citation grounding, summarization, and plagiarism detection. From this, we observed that most high-performing systems use RAG-based architectures or retrieval-augmented approaches like FiD, REALM, or HyDE. Several papers also tackled challenges like responses in source documents and detecting AI-generated or paraphrased content, directly supporting our goals of semantic search, citation extraction, and plagiarism checking. Overall, the literature confirm the technical direction of our system and offer valuable insights to guide further development of research mind.