IEEE BIG DATA 2024
December 15-18 • Washington DC, USA

IEEE International Conference on Big Data 2024

Empowering Meta-Analysis: Leveraging Large Language Models for Scientific Synthesis

Jawad Ibn Ahad, Rafeed Mohammad Sultan, Abraham Kaikobad, Fuad Rahman, Mohammad Ruhul Amin, Nabeel Mohammed, Shafin Rahman

Outline

- Introduction
  - Motivation
  - Challenges in Meta-analysis Automation
  - Existing Limitations
- Methodology Overview
- Experimental Results
- Conclusion and Future Work

Motivation

Meta-analysis is a powerful statistical approach that combines the findings of multiple studies to provide a comprehensive understanding of the same research topic.

Motivation

Meta-analysis is a powerful statistical approach that combines the findings of multiple studies to provide a comprehensive understanding of the same research topic.

- Individual studies often face limitations, such as small sample sizes or narrow focus, making it hard to draw definitive conclusions.

Motivation

Meta-analysis is a powerful statistical approach that combines the findings of multiple studies to provide a comprehensive understanding of the same research topic.

- Individual studies often face limitations, such as small sample sizes or narrow focus, making it hard to draw definitive conclusions.
- A meta-analysis paper, or **meta-article**, offers a structured analysis of numerous individual **support articles.**

Motivation

Meta-analysis is a powerful statistical approach that combines the findings of multiple studies to provide a comprehensive understanding of the same research topic.

- Individual studies often face limitations, such as small sample sizes or narrow focus, making it hard to draw definitive conclusions.
- A meta-analysis paper, or **meta-article**, offers a structured analysis of numerous individual **support articles.**

Key Concepts

- Meta-analysis is a structured synthesis of research documents.

Motivation

Meta-analysis is a powerful statistical approach that combines the findings of multiple studies to provide a comprehensive understanding of the same research topic.

- Individual studies often face limitations, such as small sample sizes or narrow focus, making it hard to draw definitive conclusions.
- A meta-analysis paper, or **meta-article**, offers a structured analysis of numerous individual **support articles.**

Key Concepts

- Meta-analysis is a structured synthesis of research documents.

Significance

- Automation strategy for multi-document meta-analysis.
- Reduces the labor-intensive aspects of meta-analysis,

Challenges in Meta-analysis Automation

Data Scarcity

- Lack of large-scale, specific meta-analysis dataset

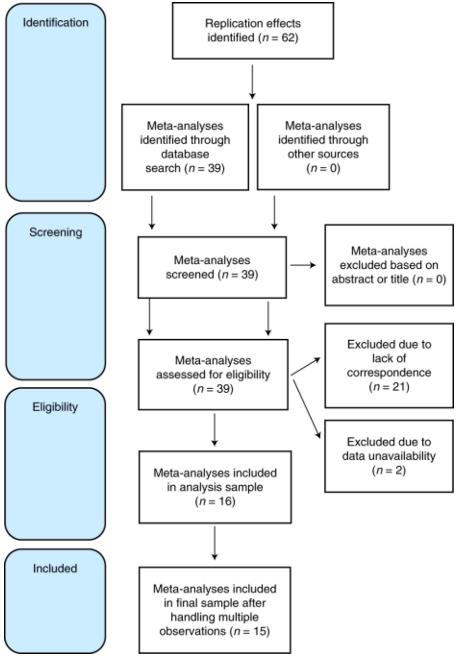
Technical Complexity

- **Handling large context length:** Meta-analysis often involves structured data derived from supporting articles, yet most LLMs operate within constrained context-length environments during fine-tuning.
- **Information loss:** Due to context length limitation, there can be information loss during training or inference.
- **Computational resource:** Fine-tuning LLMs on a large contextual corpus requires a lot of computational resources.

Existing Limitations

Traditional Approaches

- Meta-analysis presents a significant challenge to big data. The process is often labor-intensive, requiring manual extraction and analysis of data from multiple research articles, which is both time-consuming and susceptible to human error.



Existing Limitations

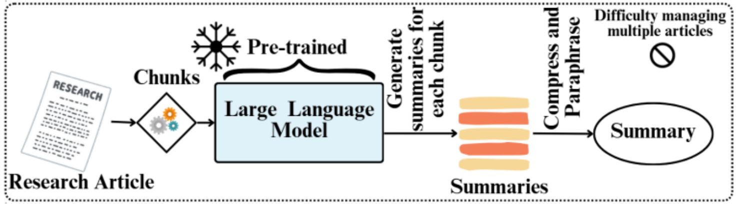
Traditional Approaches

- Meta-analysis presents a significant challenge to big data. The process is often labor-intensive, requiring manual extraction and analysis of data from multiple research articles, which is both time-consuming and susceptible to human error.

Utilizing Large language models

- **LLMs as summarizers:** Large Language models have proven their capabilities for extracting and summarizing critical context. Yet they face troubles with structured synthesis.

Paraphraser-base approach



Existing Limitations

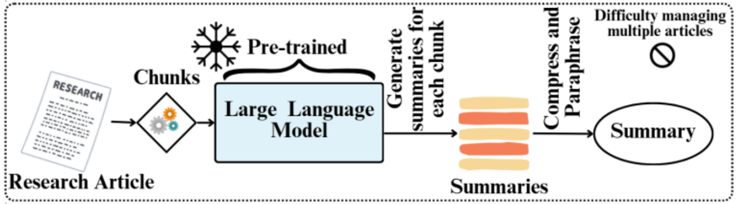
Traditional Approaches

- Meta-analysis presents a significant challenge to big data. The process is often labor-intensive, requiring manual extraction and analysis of data from multiple research articles, which is both time-consuming and susceptible to human error.

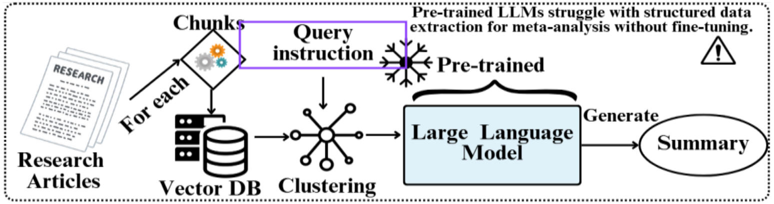
Utilizing Large language models

- **LLMs as summarizers:** Large Language models have proven their capabilities for extracting and summarizing critical context. Yet they face troubles with structured synthesis.
- **Strategy for handling large-scale data:** No approach was mentioned for handling large-context data in leveraging context-length-limited LLMs.

Paraphraser-base approach



Retrieval augmentation generation-based approach



Existing Limitations

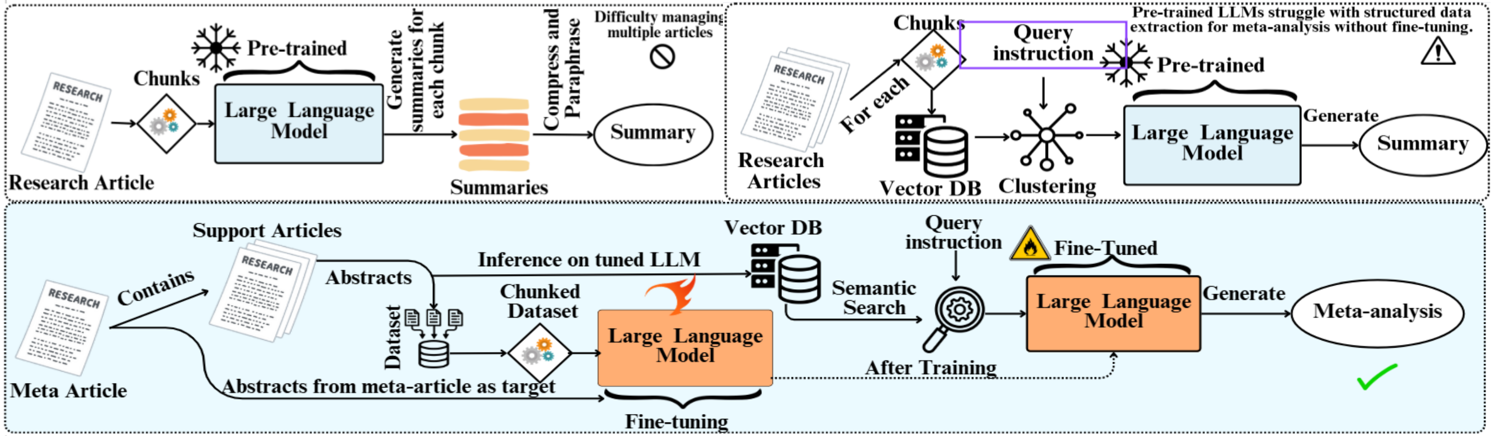
Traditional Approaches

- Meta-analysis presents a significant challenge to big data. The process is often labor-intensive, requiring manual extraction and analysis of data from multiple research articles, which is both time-consuming and susceptible to human error.

Utilizing Large language models

- **LLMs as summarizers:** Large Language models have proven their capabilities for extracting and summarizing critical context. Yet they face troubles with structured synthesis.
- **Strategy for handling large-scale data:** No approach was mentioned for handling large-context data in leveraging context-length-limited LLMs.

Paraphraser-base approach **Retrieval augmentation generation-based approach**



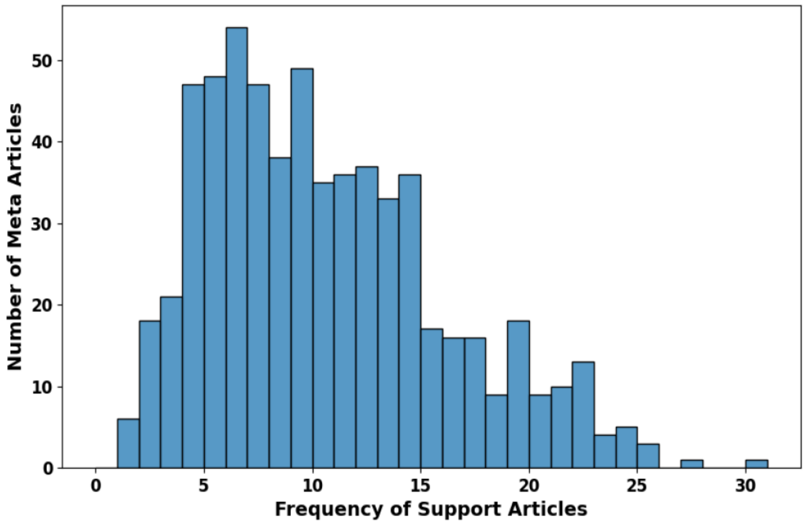
Outline

- Introduction
- Methodology Overview
  - MAD: Meta-Analysis Dataset
  - Chunk-Based Processing of Support Articles
  - Meta-analysis generation system
- Experimental Results
- Conclusion and Future Work

MAD: Meta-Analysis Dataset

MAD Statistics	Actual	Chunked
Min. input (S^j) context length	733	1005
Max. input (S^j) context length	32767	2000
Avg. input (S^j) context length	16890.22	1542.32
Min. labels (y^j) context length	104	104
Max. labels (y^j) context length	2492	2492
Avg. labels (y^j) context length	1446.45	1446.45
Total Instances	625	7447
Human Evaluators Details		
Total no. of evaluators	13	
No. of female evaluators	4	
No. of male evaluators	9	
Avg. age	23	
Profession	Student, Engineer	
Education Background	Undergraduate	

Detailed statistics of the actual and processed dataset MAD, along with the demographic profile of Human Evaluators.



Distribution of Supporting Articles in Meta-Articles in the dataset MAD.

Dataset Details

- **Number of Meta-articles:** 625
- **Number of Support articles:** 6344

Source: ScienceDirect, PubMed

Domain: Medical Science

Chunk-Based Processing of Support Articles

To manage the input size and improve model performance, we divide the support articles set, $S^j = \left\{v_1^j, v_2^j, v_3^j, \dots v_{|S^j|}^j\right\}$ into multiple smaller overlapping chunks.

Chunking of S^j into (k) possibly overlapping chunks is defined as:

$$C(S^j) = \left\{C_1^j, C_2^j, \dots, \quad C_k^j\right\}$$

Chunk-Based Processing of Support Articles

To manage the input size and improve model performance, we divide the support articles set, $S^j = \left\{v_1^j, v_2^j, v_3^j, \dots v_{|S^j|}^j\right\}$ into multiple smaller overlapping chunks.

Chunking of S^j into (k) possibly overlapping chunks is defined as:

$$C(S^j) = \left\{C_1^j, C_2^j, \dots, \quad C_k^j\right\}$$

where:

- $C_i^j \subseteq S^j$ for each $i \in [1, k]$
- $\bigcup_{i=1}^{k} C_j^i = S^j$ (the union of all chunks covers the entire set, though the chunks may overlap),
- $C_i^j \cap C_l^j \neq \emptyset$ for ($i \neq l$) (contents will overlap).

Chunk-Based Processing of Support Articles

To manage the input size and improve model performance, we divide the support articles set, $S^j = \left\{v_1^j, v_2^j, v_3^j, \dots v_{|S^j|}^j\right\}$ into multiple smaller overlapping chunks.

Chunking of S^j into (k) possibly overlapping chunks is defined as:

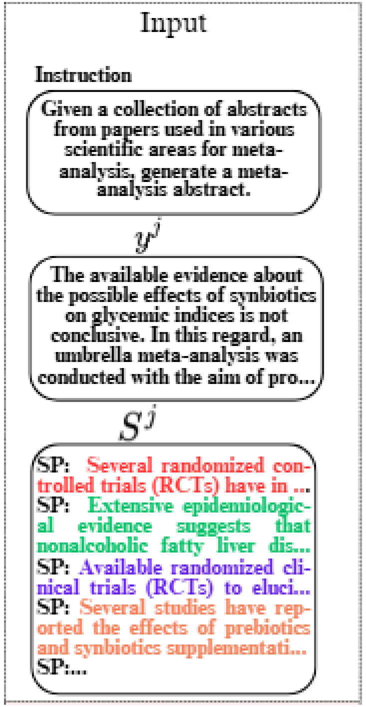
$$C\left(S^j\right) = \left\{C_1^j, C_2^j, \dots, \quad C_k^j\right\}$$

where:

- $C_i^j \subseteq S^j$ for each $i \in [1, k]$
- $\bigcup_{i=1}^{k} C_j^i = S^j$ (the union of all chunks covers the entire set, though the chunks may overlap),
- $C_i^j \cap C_l^j \neq \emptyset$ for ($i \neq l$) (contents will overlap).

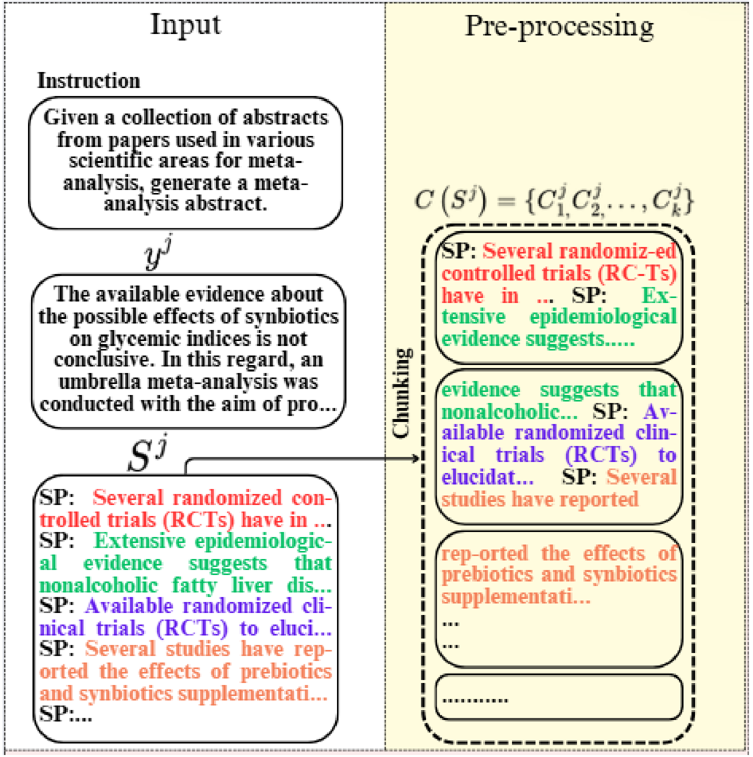
For example, suppose the set of support article abstracts $S^j = \left\{v_1^j, v_2^j, v_3^j, v_4^j, v_5^j\right\}$ is divided into three overlapping chunks. The chunking is as follows: $C\left(S^j\right) = \{C_1^j, C_2^j, C_3^j\}$ where: $C_1^j = \left\{v_1^j, v_2^j\right\}$, $C_2^j = \left\{(portion\ of) v_2^j, v_3^j, v_4^j\right\}$, $C_3^j = \left\{(part\ of) v_4^j, v_5^j\right\}$. In this example, v_2^j and v_4^j overlap in C_2^j and C_3^j.

Meta-analysis generation system



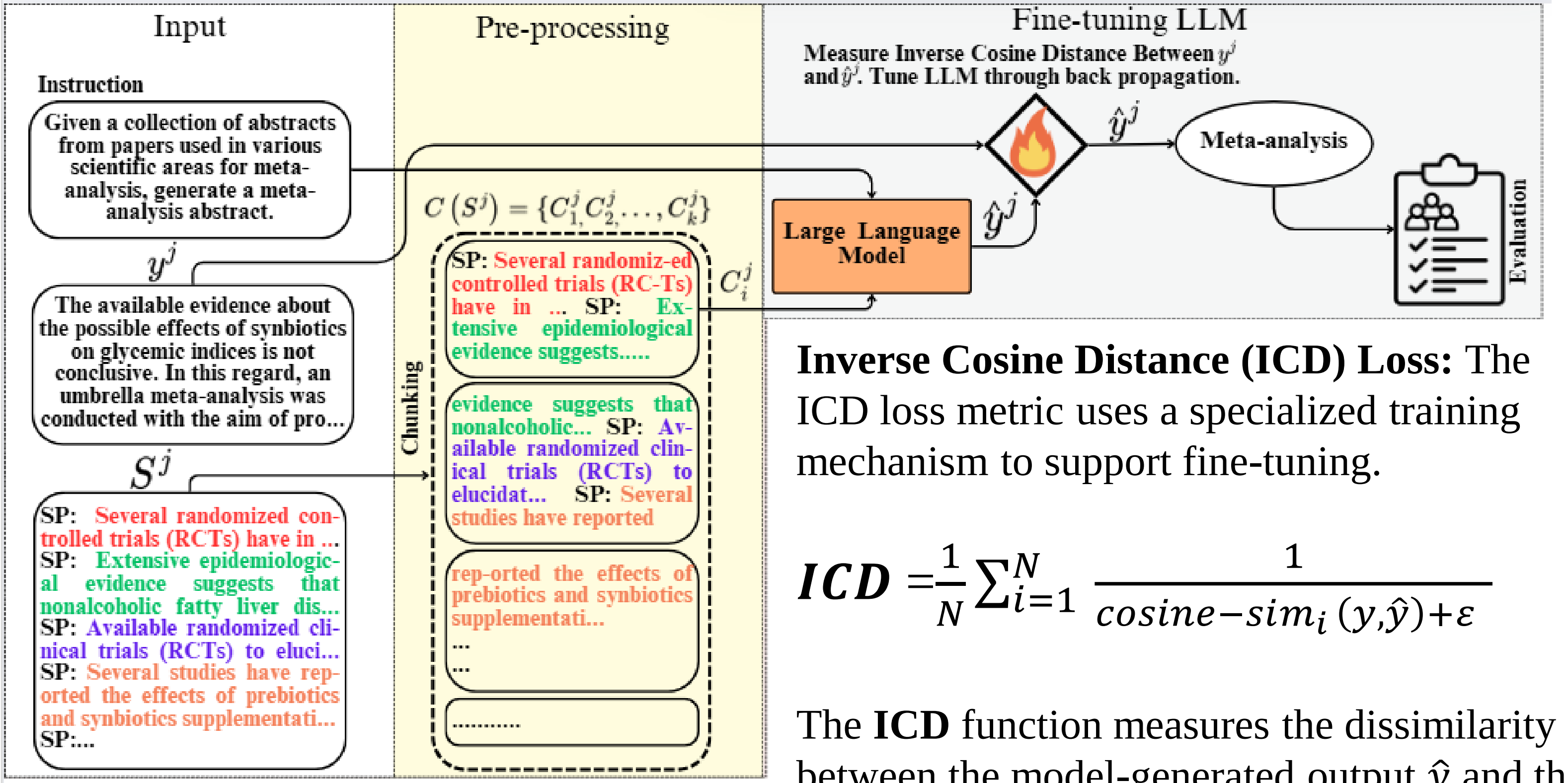
Input contains instruction and Support articles' (S^j) abstracts. For inference, there will be a query. Here **"SP:"** refers to an abstract of the support article (S^j).

Meta-analysis generation system



Support articles’ (S^j) undergo chunk-based pre-processing, producing chunks $C_i^j \subseteq S^j$. These chunks are used to fine-tune the LLMs for predicting meta-analysis abstracts (y^j), with the **ICD** loss guiding the fine-tuning process. Model performance is assessed through human evaluation of the relevancy of generated meta-analysis abstracts, $\hat{y}^j$ by fine-tuned LLMs.

Meta-analysis generation system

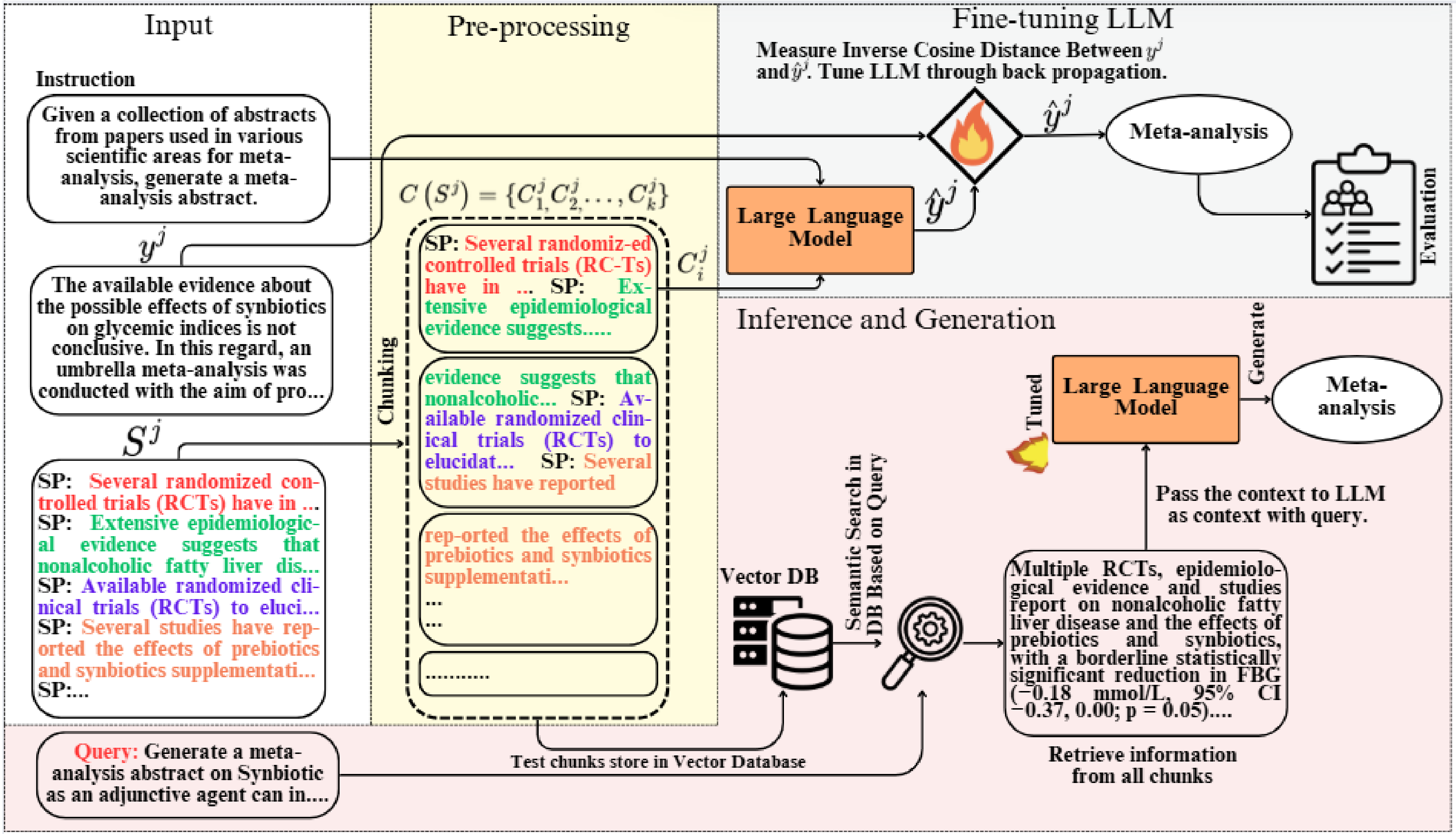


Inverse Cosine Distance (ICD) Loss: The ICD loss metric uses a specialized training mechanism to support fine-tuning.

$$ICD = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{cosine{-}sim_i\,(y, \hat{y}) + \varepsilon}$$

The **ICD** function measures the dissimilarity between the model-generated output $\hat{y}$ and the ground truth y vectors,

Meta-analysis generation system



Meta-analysis generation system

Fine-tuning LLMs

- Fine-tuning LLMs (e.g., Llama-2, Mistral) on dataset: MAD
- Overlapping segments for effective learning.
- Introduction of Inverse Cosine Distance (ICD) as a loss metric.

Inference

- Integration of Retrieval-Augmented Generation (RAG).
- Prompt engineering for targeted synthesis.

Outline

- ➢Introduction
- ➢Methodology Overview
- ➢Experimental Results
  - ❑Setup
  - ❑Results and Analysis
  - ❑Discussion
- ➢Conclusion and Future Work

Setup

➢ Dataset Details

- **Chunking:** The dataset was chunked using RecursiveSpliter.
- **Split:** 400 samples used in training, 175 in validating, and 50 in testing. After chunking, the training sample became 3659.

➢ Fine-Tuning

- Llama-2 (7B), Mistral-v0.1 (7B).
- Hardware: NVIDIA Tesla T4 GPUs.
- **Framework:** PyTorch.

➢ Evaluation

- BLEU, ROUGE, Cosine Similarity, Human Evaluation.

Results and Analysis

Human evaluation is done on the generated meta-analysis abstract by fine-tuned and non-fine-tuned LLMs, following the criteria **REL**: *Relevant*, **SWR**: *Somewhat-Relevant*, and **IRL**: *Irrelevant*. System-level metrics BLEU and ROUGE are used to identify when a human evaluator. In the end, the generated meta-analysis abstract using the RAG approach is evaluated by measuring the generated abstract's similarity with the ground truth (**SWGT**).

Models	Generated Meta-analysis Abstract							Generated Context With RAG
	Human Evaluation (%)			Relevant		Irrelevant		
	REL ↑	SWR ↓	IRL ↓	BLEU ↑	ROUGE ↑	BLEU ↑	ROUGE ↑	SWGT(%) ↑
Llama-2 7B	83.5	11.94	4.56	19.12	38.02	8.01	17.11	81.25
Llama-2 7B FT (Ours)	85.4	12.7	**1.9**	23.01	39.15	7.56	16.01	**84.32**
Mistral-v0.1 7B	80.5	14.1	5.13	22.42	39.41	**9.01**	**19.11**	77.19
Mistral-v0.1 7B FT (Ours)	**87.6**	**10.4**	2.1	**25.46**	**43.22**	7.01	17.42	83.23

Fine-Tuned LLMs:

Best model: Mistral-v0.1 7B FT (Ours).

- Relevance: 87.6%.
- Irrelevance reduction: 4.56% to 1.9%.
- BLEU and ROUGE: Consistent improvement

Key Insights:

- Fine-tuning enables models to capture contextual nuances.
- RAG improves synthesis quality.

Results and Analysis

Method	Models	Open-i[1]		writer_summaries[2]		CL-SciSumm[3]	
		BLEU ↑	ROUGE ↑	BLEU ↑	ROUGE ↑	BLEU ↑	ROUGE ↑
Established[4]	GPT-4 with ICL [1]	46.0	68.2	–	–	–	–
	InstructGPT davinci v2 [2]	–	–	–	48	–	–
	GCN Hybrid [3]	–	–	–	–	–	33.88
Context length restricted LLMs							
Pre-trained	Falcon 7B [4]	0.19	3.17	0.76	5.19	0.71	2.21
Pre-trained	Gemma 7B [5]	2.13	8.81	4.47	30.28	2.44	20.78
Pre-trained	Orca-2 7B [6]	3.53	8.36	4.29	22.51	2.86	15.55
Pre-trained	StableLM-Base-Alpha 7B [7]	2.01	2.45	3.55	15.36	2.17	16.58
Pre-trained	Llama-2 7B [8]	3.88	10.28	5.21	31.51	3.01	22.84
Pre-trained	Mistral-v0.1 7B[9]	1.21	6.57	6.32	31.36	1.05	22.55
Ours	Llama-2 7B FT	10.14	27.39	12.66	33.26	7.15	25.22
Ours	Mistral-v0.1 7B FT	**12.42**	**31.57**	**14.56**	**35.56**	**8.38**	**27.29**

[1] **Open-i:** Medical radiological dataset. We generated summaries from 100 samples.
[2] **writer_summaries:** article summarization dataset, evaluated our models on 120 samples.
[3] **CL-SciSumm:** Large corpus dataset containing scientific article data. We evaluated 20 samples. Chunking the samples was required, as the context lengths are larger than allowable.
[4] **Established** A well-established method was proven by the given papers for these three specific datasets. The BLEU and ROUGE scores given by those studies can't be achieved with our investigated LLMs, as their established methodology includes fine-tuning and evaluating the given large dataset.

[1] "Adapted large language models can outperform medical experts in clinical text summarization,"
[2] "Benchmarking large language models for news summarization,"
[3] "Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks"
[4] "The falcon series of open language models"
[5] "Gemma: Open models based on gemini research and technology"
[6] "Orca 2: Teaching small language models how to reason"
[7] "StableLM Alpha v2 Models"
[8] "Llama2: Open foundation and fine-tuned chat models"
[9] "Mistral 7b"

Results and Analysis

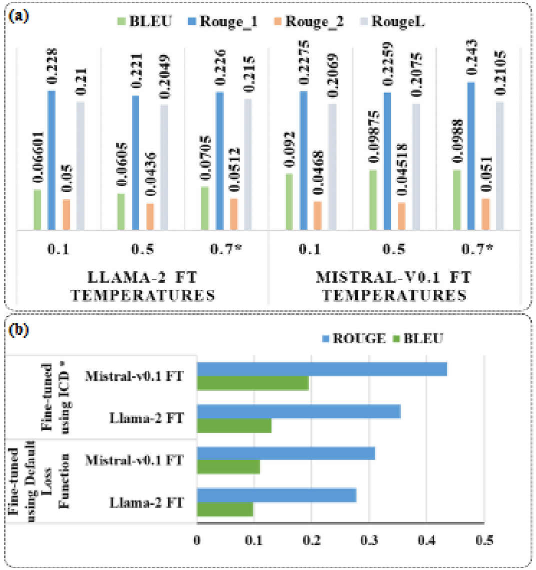
Prompt: *Given a collection of abstracts from papers used in various medical fields for meta-analysis, generate a meta-analysis abstract. Summarize the key findings and provide numerical values or statistical information for specific observations that are commonly reported in the provided abstracts.*

Example 1: Input (Support article's abstract, S^j): **SP:** Several randomized controlled trials (RCTs) have investigated the use of probiotic/synbiotic in PCOS patients,... **SP:** The effect of synbiotic supplementation on glycemic status in pregnant women remained controversial and ... **SP:** Probiotic and synbiotic products are being widely used by a large number of patients and clinicians; however, effects on cardiometabolic indices ... **SP:** This study aimed to review randomized clinical trials (RCTs) to clarify the effects of pro-/synbiotic supplementation ...fasting blood sugar (-2.52 mg/dl, 95% confidence interval (CI): -4.10 to -0.95), insulin (-2.27-µIU/mL, 95% CI: -3.40 to -1.14),...	**Generated Meta-analysis abstract, $\hat{y}^j$ (Similarity with ground truth [48]: 82.40%):** This meta-analysis abstract summarizes the findings of several randomized controlled trials (RCTs) on the effect of synbiotic intake on metabolic profiles in patients with diabetes. Heterogeneity was measured with a Q-test and the I^2 statistics. The results of the meta-analysis suggest that synbiotic intake can be useful as an adjunctive agent in the management of hyperglycemia in adults. The study pooled data by using the fix or random-effect model based on the results. A significant reduction in blood glucose levels was observed in patients who took synbiotic supplements compared to those who took placebo with statistical significance ($p < 0.05$). Additionally, a significant decrease in glycated hemoglobin (HbA1c) levels was observed in patients who took synbiotic supplements compared to placebo ($p < 0.05$). After taking synbiotic supplements, there were significant reductions in fasting blood sugar (FBS) (ES = -0.40, 95% CI: -0.64, -0.15; p = 0.002, I2 = 69.2 %, p ¡ 0.001), insulin (ES = -1.58; 95% CI: -0.79, -0.30, p ¡ 0.001, I2 = 48.4%, p = 0.042), and HOMA-IR (Homeostatic Model Assessment of Insulin Resistance) (ES = -0.55; 95% CI: -0.79, -0.30, p ¡ 0.001, I2 = 48.4%, p ¡ 0.001). Overall, the findings of this meta-analysis suggest that synbiotic intake can be an effective adjunctive agent in the management of hyperglycemia in adults.
Example 2: Input (Support article's abstract, S^j): **SP:** As the global epidemic continues to spread, countries have tappe...**SP**:Introduction: Integrated Chinese and Western medicine (integr...**SP**: A simple, efficient, and environmentally friendly electro-Fen...**SP**: Currently, coronavirus disease 2019 (COVID-19), which can lead to... **SP**: Background: Until now, there is no clinically approved spe...**SP**: Chinese medicine (CM) has been used to treat Novel Coronavi...**SP**:Integration of Chinese medical drugs (CMD) and we...**SP**:This review aims to evaluate the supportive effe...**SP**:We systematically studied the passivation process of 6082 aluminium alloy under the bending stress... **SP**: Coronavirus disease 2019 (COVID-19) is an eme...**SP**:There is currently no drug or therapy that cures COVID-19, a highly contagious and... **SP**: The outbreak of coronavirus disease 2019...**SP**: Coronavirus disease 2019 (COVID-19) has eme...**SP**:Background: The coronavirus disease 2019 (COVID-19) pandemic...	**Generated Meta-analysis abstract, $\hat{y}^j$ (Similarity with ground truth [49]: 85.73%):** In this meta-analysis, we evaluated the effectiveness of traditional Chinese medicine (TCM) in treating COVID-19. A comprehensive search was conducted on a variety of electronic databases to identify studies reporting TCM treatment for COVID-19. The studies were screened according to inclusion and exclusion criteria, and a total of 10 studies were included in the analysis. These studies had a total of 1,880 participants who received TCM treatment for COVID-19. The analysis found that TCM had significant positive effects on the symptomatic relief of COVID-19. The pooled effect size for symptomatic relief was 0.67 (95% CI: 0.56 to 0.79), indicating a moderate effect size. The studies also reported on the impact of TCM on the severity of COVID-19. The pooled effect size for severity reduction was 0.51 (95% CI: 0.43 to 0.60), indicating a moderate effect size. In addition, the studies also reported on the impact of TCM on the mortality rate of COVID-19. The pooled effect size for mortality reduction was 0.37 (95% CI: 0.21 to 0.55), indicating a small but statistically significant effect size. Overall, this meta-analysis suggests that TCM may be an effective treatment for COVID-19, and further studies are needed to confirm these findings. It is important to note that this meta-analysis had several limitations, including the small number of studies, heterogeneity of TCM treatments, and variable outcomes. Out of the 89 outcome indicators that were assessed, 8 were found to be of moderate quality, 39 to be of low quality, and 41 to be of very low quality. In conclusion, TCM may be an effective treatment for COVID-19, with significant positive effects on symptomatic relief and severity reduction. However, more research is needed to confirm these findings and to establish the optimal TCM treatment for COVID-19.

Discussion

Ablation Study

- ➢ **Impact of ICD loss**
  - It is observed that ICD loss works better than the default loss during training.
- ➢ **Temperature**
  - It is found that the temperature at 0.7, works better than 0.1 and 0.5.



Prompt	Evaluation Metric	Llama-2	Mistral	Llama-2 Ours	Mistral Ours
Given a collection of abstracts from papers used in various medical fields for meta-analysis, generate a meta-analysis abstract. Summarize the key findings and provide numerical values or statistical information for specific observations that are commonly reported in the provided abstracts. (Prompt 1)	Relevant ↑	**83.5**	**80.5**	**85.4**	**87.6**
	Somewhat Relevant ↑	11.94	14.1	12.7	10.4
	Irrelevant ↓	4.56	5.13	1.9	2.1
There are given some abstracts of papers that are used for meta-analysis in different medical fields. Generate a meta-analysis abstract based on the given abstracts of papers. Please try to provide numerical values for any specific findings that were used in most of the abstracts. (Prompt 2)	Relevant ↑	69	78.39	72.4	82.8
	Somewhat Relevant ↑	12.7	16.1	20.5	14.1
	Irrelevant ↓	7.69	5.51	7.1	3.13

Discussion

➢ Key Contributions

- Comprehensive dataset for large-context meta-analysis.
- Fine-tuning methodology tailored for structured scientific synthesis.
- Demonstrated scalability and efficiency of automated meta-analysis.

➢ Limitations

- Restricted to specific context lengths.
- High computational requirements.

Outline

- Introduction
- Methodology Overview
- Experimental Results
- Conclusion and Future Work

Conclusion

Summary

- Developed an automated pipeline for **meta-analysis generation** using fine-tuned **Large Language Models (LLMs)**.
- Introduced a novel dataset (**MAD**) and methodologies like **chunking**, **RAG**, and a custom loss function (**ICD**) to enhance model performance.
- Achieved significant improvements in relevance and reduced irrelevance in meta-analysis abstracts.
- Demonstrated the potential of LLMs in synthesizing large-scale scientific data efficiently and accurately.

Impact

- **Accelerates Research:** Streamlines meta-analysis, reducing time and effort for researchers.
- **Improves Accuracy:** Reduces errors in meta-analysis synthesis compared to manual methods.
- **Broad Applications:** Can benefit fields like healthcare, environmental science, and education by enabling data-driven decision-making.
- **Pioneers New Approaches:** Sets a foundation for automating complex scientific processes with AI.
- **Scalable Solution:** Offers a pathway to expand into multilingual and cross-domain applications in the future.

Future Work

In conclusion, our research demonstrates the potential of LLMs for automating meta-analysis. By fine-tuning these models and using tools like RAG, we've made the process faster and more reliable.

Looking ahead, we plan to:

- Expand the dataset to include more fields.
- Improve the fine-tuning process to handle larger contexts.
- Develop user-friendly tools for real-world applications, like healthcare and education.
- Combine multimodal data (eg. Images) to reproduce strategies that are followed in other studies.

Thank You

Apendix

Meta-paper: Intervention methods for improving reduced heart rate variability in patients with major depressive disorder: A systematic review and meta-analysis

Table 1
Summary of the included studies.

First author /time	Intervention groups	Sample size	Duration of intervention	HRV outcome indices	HRV measurement	Significant results Inner group
Pharmacotherapy						
Hartmann/ 2018	G1: Antidepressants (unspecific)	62	2 weeks	SDNN, RMSSD, pNN50, LF (ms2, %, nu), HF (ms2, %, nu), LF/HF, SD1, SD2, SD1/SD2	Resting-state; eyes closed; duration not mentioned	HF (%) ↑ , HF (nu)↑ , LF(nu)↓ , SD2↓, SD1/SD2↓
	G2: Health control	65				/
Koenig/ 2018	Fluoxetine or Sertraline, adolescent	12	8 weeks	HR, RMSSD, HF	Resting-state; eyes closed; 5 mins	HF↑
Chang/2018	G1: Agomelatine	56	6 weeks	Mean R-R intervals (ms), Variance (total HRV), VLF, LF, HF, LF/HF (all by Ln)	Resting-state; 5 mins	HF↑
	G2: Agomelatine+sedatives	49				LF↓, LF/HF↑
Yeh/2016	G1: Agomelatine	29	6 weeks	Mean R-R intervals (ms), Variance (total HRV), VLF, LF, HF, LF/HF (all by Ln)	Resting-state; 5 mins	HF↑
	G2: Paroxetine	27				NS
Terhardt/ 2013	G1: Mirtazapine	24	4 weeks	Rest and activation: LF/HF, TP	Supine and standing position; 5 mins each	Rest and activation TP↓
	G2: Venlafaxine	20				Rest and activation TP↓
Davidson/ 2005	G1: Venlafaxine	24	3 weeks	R-R interval changes (ΔR-R), RSA	Continuous deep breathing	ΔR-R↓, RSA ↓
	G2: Paroxetine	25				NS
Bär/2003	G1: SSRIs/SNRIs	30	B: 2–4 days C: 6–9 months	CV, RMSSD, VLF, LF, HF, LF/HF	Resting-state; 5 mins	B and C: CV↓, RMSSD↓
	G2: Health control	30				/
Rechlin/ 1994	G1: Amitriptyline	8	2 weeks	CV, RMSSD	Resting-state; 5 mins	CV↓, RMSSD↓
	G2: Doxepin	8				CV↓, RMSSD↓
	G3: Paroxetine	8				NS
	G4: Fluvoxamine	8				NS
Psychotherapy						
Caldwell/ 2018	G1: CBT/ACT+HRVB, young	10	6 weeks	SDNN, LF, HF, LF/HF	Resting-state; 5 mins	NS
	G2: CBT/ACT, young	10				NS
	G3: Healthy controls	10				NS
Dixon/2017	G1: Acceptance-based ER	36	78 min	HF	Sitting still; 5 mins	NS
	G2: Suppression-based ER	37				NS
Chien/2015	G1: CBT + breathing relaxation	43	4 weeks	LF, HF, LF/HF, SDNN	Sitting still; 5–7 min	LF↑, HF↑, LF/HF↓, SDNN↑
	G2: TAU control	46				NS
Kim/2009	G1: CBT in Forest	23	4 weeks	SDNN, RMSSD, TP, HF, LF, VLF, LF/HF, Triangular index	Relaxed sitting; 5 mins	SDNN↑, RMSSD↑, HF↑, TP↑

	G2: CBT/ACT, young	10				NS	
	G3: Healthy controls	10				NS	
Dixon/2017	G1: Acceptance-based ER	36	78 min	HF	Sitting still; 5 mins	NS	NS
	G2: Suppression-based ER	37				NS	
Chien/2015	G1: CBT + breathing relaxation	43	4 weeks	LF, HF, LF/HF, SDNN	Sitting still; 5–7 min	LF↑, HF↑, LF/HF↓, SDNN↑	G1 vs G2: LF↑, HF↑, LF/HF↓, SDNN↑
	G2: TAU control	46				NS	
Kim/2009	G1: CBT in Forest	23	4 weeks	SDNN, RMSSD, TP, HF, LF, VLF, LF/HF, Triangular index	Relaxed sitting; 5 mins	SDNN↑, RMSSD↑, HF↑, TP↑	/
	G2: CBT in hospital	19				NS	
	G3: Usual outpatient control	21				NS	

First author /time	Groups	Sample size	Duration of intervention	HRV outcome indices	HRV measurement	Significant results Inner group	 Inter groups
Physiotherapy							
Bozkurt/ 2013	ECT, resistant MDD	14	6 weeks	pNN50, RMSSD, LF, HF, LF/HF	Completely rest in bed; unclear	Non-responder (n = 7) HF↓	/
Royster/ 2012	ECT	21	12 days	RMSSD, LF, HF, LF/HF, SD1, SD2, SD1/SD2	Resting-state; 15 mins	NS	/
Ebert/2010	ECT	24	6 sessions	SDNN, LF nu (LF/[TP-VLF]), HF nu (HF/[TP-VLF])	Resting-state; 30 mins	NS	/
Nahshoni/ 2004	ECT, older adults	10	3–6 weeks	LF, HF, PD2	Resting-state; 2000 R-R	Responder (n = 7) PD2↑	/
Nahshoni/ 2001	ECT, older adults	11	3–6 weeks	TP, LF, HF, LF/HF, LF norm, HF norm	Resting-state; 2000 R-R	LF norm↓, HF norm↑, LF/HF↓	/
Iseger/2020	50% sham +50% active iTBS	15	30 days	VLF, LF, HF, LF/HF, SDNN, RMSSD (all by Ln)	189 s	Active > Sham: LF↑, HF↑, SDNN↑, RMSSD↑	/
Combined intervention							
Toni/2016	Late-life MDD G1: Sertraline	50	24 weeks	R-R interval, pNN50, RMSSD, SDNN, HF, LF, LF/HF	Resting-state; 20 mins	All participants: pNN50, RMSSD, SDNN, LF, and HF increased from baseline, and the LF/HF ratio decreased	G2 showed greater increases of pNN50, RMSSD, SDHR, SDNN, HF, and LF
	G2: Sertraline+PE	38					

(*continued on next page*)