AGORA Revisited: A Replication Study on Innovative Test Oracle Generation for Advancing the Robustness of REST APIs

SUMMARY

- Evaluate AGORA's performance in REST API test oracle generation.
- Replicate original research paper's methodology.
- Conduct experiments with diverse datasets and duplicate API request control.
- Gain valuable insights into AGORA's behavior, highlighting areas for improvement in automated API testing.

MOTIVATION

- REST APIs are integral to modern software development.
- Challenges in ensuring API robustness and reliability.
- Validate and extend findings from original AGORA research.

1. Scalability with Larger Datasets:
Investigate AGORA's performance with increasing dataset sizes.
Crucial for practical deployment in real-world scenarios.
Identify scalability limitations and strategies for improvement.

2. Impact of Duplicate Request Reduction:
Examine effects of detecting and reducing duplicate requests.
Enhance AGORA's efficiency and effectiveness.
Streamline testing processes and improve reliability of API testing outcomes.

ANALYSIS & RESULTS - RQ1

Extension of Dataset Size

Table which represents the OMDB ByIdOrTitle invariants of different dataset sizes

The dataset size	50	100	1000	10000	15000
Invariants detected	11	10	9	12	15

Evaluation of YouTube Data API with 10,000 Responses:
Insights into API Behavior with Varied Dataset Sizes

(Youtube)	Current_Invariant	Modified_Invariant
The dataset size	10000	10000
Invariants	132	139

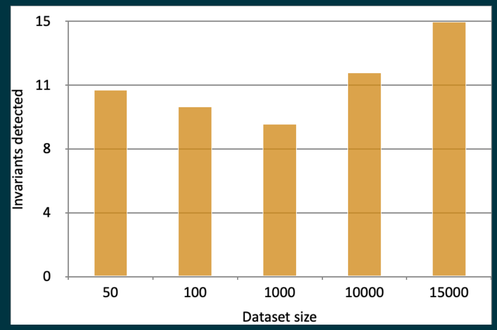
The Research Questions

- How does the performance of AGORA scale with larger datasets of API requests and responses? Are there any bottlenecks in the system as the dataset size increases?
- How does detecting and reducing duplicate input requests affect the results (if it captures new invariants?) of AGORA's API testing process? (Data deduplication using a threshold value)

RQ1 Findings

AGORA consistently captures key API behaviors and reveals deeper, more specific patterns as dataset sizes increase, demonstrating its scalability and enhanced behavioral analysis capabilities.

Larger datasets enable a more comprehensive understanding of API behavior by introducing additional invariants without increasing false positives, ensuring the precision of AGORA's test oracles.



RQ2 Findings

Despite the emergence of new invariants in the Spotify dataset following deduplication at a threshold value of 5, other datasets failed to yield desired outcomes. Notably, some endpoints within alternative datasets maintained comparable invariant counts or exhibited minimal changes.

DATA AVAILABILITY

GITHUB

https://github.com/parthcompengg/CS569-AGORA-ProjectGroup8

TRELLO

https://trello.com/b/PwYf9m1i/agora-project-group-8

ANALYSIS & RESULTS - RQ2

- Threshold Value Set: Established threshold value of 5 through iterative testing.
- Iterative Approach: Started with higher values, gradually reducing to observe changes.
- Critical Threshold Point: Significant changes observed at threshold value of 5, especially in Spotify dataset.
- Duplicate Records Limitation: Implemented measures to restrict duplicate records in dataset.
- Input Modification: Provided modified dataset and OpenAPI Specification to Beet.
- Invariants Generation: Used modified Daikon tool to produce invariants.

	Results for Existing datasets					Results for Deduplicated datasets				
	10k	1000	500	100	50	10k	1000	500	100	50
Endpoints										
Spotify										
createPlaylist	41	41	41	41	41	50	43	43	43	46
getAlbumTracks	90	84	84	83	80	90	84	84	86	80
getArtistAlbums	92	84	84	84	95	94	84	84	84	95
Amadeus Hotel										
getMultiHotelOffers	226	227	228	220	223	226	227	228	220	223
Yelp										
	142	147	149	147	130	142	147	150	147	130
Github										
createOrganizationRepository	210	210	210	210	210	210	210	206	210	210
getOrganizationRepositories	169	169	168	164	167	169	168	168	164	167
OMDb										
byIdOrTitle	32	33	33	33	33	32	33	33	33	33
bySearch	28	25	25	20	22	28	25	25	20	22

Number of unique invariants identified

CONCLUSION

Performance Insights:
- AGORA shows promise in detecting invariants and API errors.
- Performance varies based on dataset composition and threshold values.
- Threshold value of 5 significantly impacted Spotify dataset, but not others.

Future Research Directions:
- Explore strategies to enhance AGORA's robustness.
- Refine algorithms to improve performance across diverse datasets.

The Research paper: