ITRIX - an AI Enabled Solution for Orchestration of Recovery Instructions

Wanita Sherchan[1], **Garfield Vaughn**[2], **Shaila Pervin**[1], **Bryan Barone**[2]
[1] IBM Research Australia
[2] IBM Global Technology Services
wanita.sherchan@au1.ibm.com, garfield@us.ibm.com, shaila.pervin@au1.ibm.com, bryanbarone@us.ibm.com

Abstract

This paper presents ITRIX (IT Recovery Instructions eXtraction)- a solution that automates the complex manual process of creating executable recovery plans from textual documents describing IT environment disaster recovery procedure. The solution applies natural language processing and deep learning to identify textual content that contain recovery instructions and information relevant to disaster recovery. This enables rapid creation and review of recovery plans resulting in 80% reduction in time taken to generate them.

Introduction

To stay relevant and competitive, organizations depend on IT solutions to manage and support their systems and business processes. An outage for any of these systems means a loss of revenue, disruption of productivity, and a negative impact to client satisfaction and the company's reputation. A critical insurance policy most organizations have invested in is a Disaster Recovery (DR) solution (Tamimi, Dawood, and Sadaqa 2019). The Disaster Recovery as a Service (DRaaS) market is estimated to be $4B (USD) and expected to reach over $20 Billion in the next three years. IBM is currently one of the leading service providers in this space. Orchestrated disaster recovery is a key area where providers are building custom integrated solutions to differentiate themselves from other competitors. To transition from a manually executed recovery to an automated/orchestrated recovery requires translating recovery instructions from document files (Word, PDF, Spreadsheet) to machine readable instructions. Today this is a highly manual process that requires subject matter experts (SMEs) to read the instructions, identify recovery steps, then re-key or copy/paste them into a spreadsheet which is then used to build runbook(s)- a compilation of routine procedures and operations that the system administrator/operator needs to carry out for IT environment recovery. This dependency on technical experts with deep and wide-ranging knowledge gained over years is a significant challenge and has implications on the scalability of DRaaS delivery, the capability to meet increasing demands and the need for improved client satisfaction. In this paper we present IT Recovery Instructions eXtraction (ITRIX), an AI supported solution, that has been developed and proven to significantly reduce the time it takes to identify and extract recovery instructions from unstructured data contained in Technical Recovery Procedure (TRP) documents. Deep learning based natural language processing (NLP) has been applied in many application domains (Otter, Medina, and Kalita 2020). This paper presents a novel application in the domain of disaster recovery orchestration.

Recovery Instructions Extraction Modelling

The ITRIX solution utilises a combination of NLP models to automatically create an executable list of instructions for IT system recovery in the event of an outage. It employs a Long Short-Term Memory (LSTM) model to extract recovery instructions/actions and information from a TRP document. Out of the sentences that are determined to be recovery actions, a rules based model is applied to identify whether the recovery action is a critical recovery action. For each recovery action identified to be a critical action, system added instructions/actions are inserted where the user is instructed to check the status before and/or after executing a critical action. Some examples of critical actions are shutdown and startup actions, updates or alteration of configurations and database operations. This step prevents unintended consequences of performing or executing a critical action without verifying the current status of the system. Some of these critical actions can be disastrous if executed without these checks and if re executed. Checking critical actions and inserting system added actions aids graceful disaster recovery.

The LSTM model was developed using AllenNLP (Gardner et al. 2017) and cc-models library (Sherchan et al. 2020) and deployed using Flask[1], Swagger[2] and Docker[3]. The model achieved an F1 score of 0.93 on test data sets. Model performance was further validated with qualitative evaluation by 5 SMEs in a pilot over a 1 month period. The pilot demonstrated that ITRIX reduced the time taken for the creation of runbooks by 80%. Since then, the solution has been deployed and is being used by a broader team of SMEs.

Another aspect of the solution is that a single TRP document may contain recovery instructions for a number of

[1]http://flask.pocoo.org/
[2]https://swagger.io/
[3]https://www.docker.com/

scenarios. For example, instruction sets for a switchover or failover scenario (i.e., bring down the production systems and bring up the back up systems) and another set for switchback or failback (i.e., bring down the back up systems and bring up the production systems). Finally, in some cases, a test scenario may be defined where the instructions are for testing the failover procedure without actually shutting down the production systems. The use of NLP models helps in identifying different sets of recovery scenarios present and any missing scenarios. For example, if switchover/failover use case exists then switchback/failback should also exist and vice versa whereas a failover test scenario may exist on its own. This makes it possible to generate instruction sets for all scenarios from a single TRP document.

Solution Architecture

Figure 1 shows the document processing and user interaction flow in ITRIX. The process starts with uploading the client supplied recovery procedure document (TRP). The document is parsed by the Document Parser module to extract all textual content. The textual content is processed by the LSTM model described above (Action Classifier Model) to extract recovery instruction sentences and information sentences related to recovery. The recovery instructions are processed by the Critical Actions Model to identify whether any of them is a critical instruction and then system generated checks/instructions are inserted as appropriate before and/or after the critical instructions. The set of instructions are then split into scenarios by the Scenario Classifier Model. The instructions set per scenario is then available for the user to review as shown in Figure 2. These action/instruction items can be iterated over with multiple parties using a workflow. Users can also provide feedback to help improve the AI models. Further, users can add configuration details to the action/instruction items such as user names and passwords, host names, ip addresses, database types, etc. When the list of action items are finalised by the user and appropriate configuration information has been populated, these instructions will be mapped to executable code by the Recovery Actions Library (RAL) Classifier Model. The set of executable code is then ready to be tested and run by the users as the Executable Recovery Plan for that environment.

Conclusion

In this paper we have described ITRIX - a novel solution that applies natural language processing, machine learning and deep learning techniques to extract relevant recovery instructions from a textual document. ITRIX is a web application that helps improve the delivery of Resiliency Orchestration (RO). With the help of the AI assisted solution, RO consultants and specialists are able to easily extract recovery actions from client recovery procedure documents which is currently a fully manual operation. Additionally, in a single interface users can build out the IT recovery procedure of the client's infrastructure as well as applications. ITRIX enables rapid creation and review of recovery plans with initial trials demonstrating an 80% reduction in time taken for recovery plans generation.

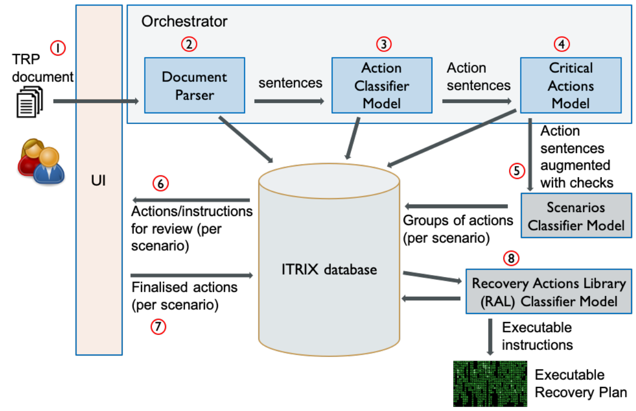

Figure 1: ITRIX solution architecture

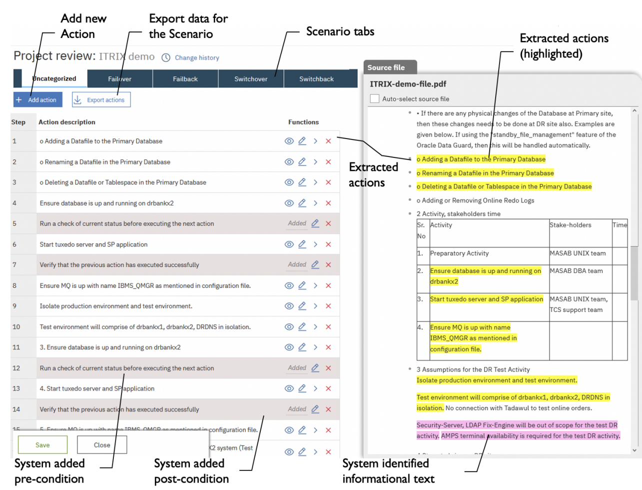

Figure 2: ITRIX UI depicting the SME view of recovery instructions extracted by the AI solution

References

Gardner, M.; Grus, J.; Neumann, M.; Tafjord, O.; Dasigi, P.; Liu, N. F.; Peters, M.; Schmitz, M.; and Zettlemoyer, L. S. 2017. AllenNLP: A Deep Semantic Natural Language Processing Platform.

Otter, D. W.; Medina, J. R.; and Kalita, J. K. 2020. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems* 1–21. doi:10.1109/TNNLS.2020.2979670.

Sherchan, W.; Chen, S.; Harris, S.; Alam, N.; Tran, K.-N.; and Butler, C. 2020. Cognitive Compliance: Assessing Regulatory Risk in Financial Advice Documents. *Proceedings of the AAAI Conference on Artificial Intelligence* 34: 13636–13637. doi:10.1609/aaai.v34i09.7105.

Tamimi, A. A.; Dawood, R.; and Sadaqa, L. 2019. Disaster Recovery Techniques in Cloud Computing. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 845–850.