

A Gold Standard Business Process Model Collection

By

RaoFaizan Ali

A dissertation submitted in fulfillment of
The requirements for the degree of Master of Science in Computer Science

Principal Supervisor:Dr.AdnanAbid

School of Science and Technology
Computer Science Faculty
University Of Management and Technology
Lahore, Pakistan

September 2015

Dedication

I dedicate my research work to my beloved teachers and role models Dr. KhurramShahzad, Dr Adnan Abid and DrAdeelNawab, their help and devotion made this happen. They supported me on every difficult node of this work. Because of them I learned how to research and how to prove a technique with strong arguments. They encourage me to believe in myself.

Keywords

Business process management, business process model, process model collection, Process Model Repository, Light Revision, Near copy, Heavy Revision, Gold Standard, CloneMatching, Filtration, Process Models Similarity, SAP, APROMORE, Synthetic Models.

ABSTRACT

Business Process modeling enables an organization to analyze their business operations and the change comes into those operations with the passage of time. There are too many processes in an organization; for managing those processes organizations maintain repositories of process models. However process models repositories are build in different stages and by different people in the large organizations so with that case repository must have data duplication and noisy data. For managing that data there must be some automated technique required.

To managing those noisy collections researchers are trying to build repositories which are advanced and manageable easily. For making those kind of repositories there are some advanced features required .these features are classified into four types, evaluation, comparison, management and presentation. Whereas evaluation covers the analysis of quality, correctness and performance analysis .comparison considers similarity between models and pattern based analysis. Management describes about how to control extensions, individualization and reference based completion of the repository. Last and most useful feature is presentation of the process models presentation relates to useful generalization of processes and secondary notations (size,color,etc).

As people have tried to invent several automatic techniques for matching, it is very important to rigorously evaluate these approaches. If the approaches are not evaluated rigorously, there are chances that the approaches our work for process models of specific category, such as large process models or dense process models. Also, one would like to be able to say which are better, which worse, and also which words, or varieties of issues, present particular problems to which algorithms. In order to evaluate approaches to process matching/detection it is useful to have access to a collection of models that containing examples of the types of models/models of different features that we aim to identify. To assist in this study it was necessary to create a comparable corpus consisting of a selection of process models that have specific features.

The purpose of this research was to develop a gold standard collection of process models, classification of process models, develop queries on defined classification and matching those

queries to determine validity of collection. In our research work, we have built a collection of process models which are in BPMN 2.0. After creating the corpus of process models we classified them. The task of classification will be done based on structural Metrics like size, type etc. A set of queries will be made from each class then we performed query matching on the collection of process models. For process models matching Manual approaches of similarity computation used.

Contents

Dedication.....	2
Keywords.....	3
ABSTRACT.....	5
List of Figures	11
List of Tables	12
1. Introduction.....	Error! Bookmark not defined.
1.2 BPM Introduction	Error! Bookmark not defined.
1.3 Repository Technology for Process Models	Error! Bookmark not defined.
1.4 Need for the Dataset, and Usability of the Dataset	Error! Bookmark not defined.
1.4.1 Clone Detection:	Error! Bookmark not defined.
1.4.2 Evaluation of Process Models:.....	Error! Bookmark not defined.
1.4.3 Filtering:.....	Error! Bookmark not defined.
1.4.4 Clever Design:	Error! Bookmark not defined.
1.4.5 Presentation:.....	Error! Bookmark not defined.
1.5 Research Question	Error! Bookmark not defined.
1.6 Research Objective	Error! Bookmark not defined.
1.7 Research Methodology	Error! Bookmark not defined.
1.8 Thesis Structure	Error! Bookmark not defined.
2. Literature Review.....	Error! Bookmark not defined.
2.1 Existing Repositories and Their Features	Error! Bookmark not defined.
2.1.1 MIT process handbook.....	Error! Bookmark not defined.
2.1.2 Advanced Process Model Repository (APROMORE)	Error! Bookmark not defined.
2.1.2.1 Evaluation	Error! Bookmark not defined.
2.1.2.2 Filtering.....	Error! Bookmark not defined.
2.1.2.3 Design	Error! Bookmark not defined.
2.1.2.4 Presentation.....	Error! Bookmark not defined.
2.1.3 The Repository for Integrated Process Management	Error! Bookmark not defined.

2.2	The Existing Collections	Error! Bookmark not defined.
2.2.1	APROMORE Data Collection	Error! Bookmark not defined.
2.2.2	MIT Handbook Data Collection	Error! Bookmark not defined.
2.2.3	SAP Collection.....	Error! Bookmark not defined.
2.2.4	Synthetic Models	Error! Bookmark not defined.
2.3	Limitations of Existing Collections	Error! Bookmark not defined.
2.3.1	APROMORE Limitations:.....	Error! Bookmark not defined.
2.3.2	MIT Handbook Limitations:	Error! Bookmark not defined.
2.3.3	C.SAP Collection Limitations.....	Error! Bookmark not defined.
2.3.4	D. Synthetic Models Limitations	Error! Bookmark not defined.
3	The Process Model Collection	Error! Bookmark not defined.
3.1	The Process of Dataset Generation	Error! Bookmark not defined.
3.1.1	The Dataset Team	Error! Bookmark not defined.
3.1.2	Collecting the Models	Error! Bookmark not defined.
3.1.3	Modeling Process in Signivio	Error! Bookmark not defined.
3.1.4	Enriching the Models	Error! Bookmark not defined.
3.1.4.1	Modeling Guidelines Implementation In our Collection.	Error! Bookmark not defined.
3.1.4.2	Use One Start and One End Event.....	Error! Bookmark not defined.
3.1.4.3	Each Split Gateway Must have its Subsequent Join	Error! Bookmark not defined.
3.1.4.4	Each Corresponding Gateway Type should be Same	Error! Bookmark not defined.
3.1.4.5	Arcs and Gateways Should Not be Labeled.....	Error! Bookmark not defined.
3.1.4.6	Each Activity Must Have One Incoming and One Outgoing Edge	Error! Bookmark not defined.
3.1.4.7	Loops Should be Avoided.....	Error! Bookmark not defined.
3.1.4.8	Each Activity Must be Labeled.....	Error! Bookmark not defined.
3.1.4.9	Each Arc Must Have its Target or Source	Error! Bookmark not defined.
3.1.4.10	Use Pools and Lanes	Error! Bookmark not defined.
3.1.4.11	Use Only 1 Pool and at Least One Lane	Error! Bookmark not defined.
3.1.4.12	Arcs/Activities/Gateways Should Not Overlap	Error! Bookmark not defined.
3.1.5	Collection of Models and Their Analysis Using Structural Metrics.....	Error! Bookmark not defined.
3.1.5.1	Size.....	Error! Bookmark not defined.
3.1.5.2	Diameter.....	Error! Bookmark not defined.

3.1.5.3	Density	Error! Bookmark not defined.
3.1.5.4	Coefficient of Connectivity.....	Error! Bookmark not defined.
3.1.5.5	Average Degree of Connectors	Error! Bookmark not defined.
3.1.5.6	Maximum Degree of Connectors.....	Error! Bookmark not defined.
3.1.5.7	Separability Ratio.....	Error! Bookmark not defined.
3.1.5.8	Sequentiality Ratio.....	Error! Bookmark not defined.
3.1.5.9	Connector Mismatch.....	Error! Bookmark not defined.
3.1.5.10	Cyclicity.....	Error! Bookmark not defined.
3.1.5.11	Token Split.....	Error! Bookmark not defined.
3.1.5.12	S (N).....	Error! Bookmark not defined.
3.1.5.13	S (E)	Error! Bookmark not defined.
3.1.5.14	S (ES).....	Error! Bookmark not defined.
3.1.5.15	S (EE).....	Error! Bookmark not defined.
3.1.5.16	S (F)	Error! Bookmark not defined.
3.1.5.17	S (C).....	Error! Bookmark not defined.
3.1.5.18	S (C) AND	Error! Bookmark not defined.
3.1.5.19	S (C) XOR.....	Error! Bookmark not defined.
3.1.5.20	S (C) OR	Error! Bookmark not defined.
3.1.5.21	S (J) AND	Error! Bookmark not defined.
3.1.5.22	S (J) XOR.....	Error! Bookmark not defined.
3.1.5.23	S (J) OR.....	Error! Bookmark not defined.
3.1.5.24	S (S) AND.....	Error! Bookmark not defined.
3.1.5.25	S (S) XOR.....	Error! Bookmark not defined.
3.1.5.26	S (S) OR.....	Error! Bookmark not defined.
3.1.6	Process Models Conversion to JSON Format.....	Error! Bookmark not defined.
3.2	Observations about Collection	Error! Bookmark not defined.
3.3	Summary of Models collected	Error! Bookmark not defined.
3.4	Process Models Revisions.....	Error! Bookmark not defined.
3.4.1	Near copy.....	Error! Bookmark not defined.
3.4.2	Light Revision.....	Error! Bookmark not defined.
3.4.3	Heavy Revision.....	Error! Bookmark not defined.

3.4.5	Summary of the Models and Their Revision as Well as Their Discussion...	Error! Bookmark not defined.
4	Analysis of the Collection.....	Error! Bookmark not defined.
4.1	The Metrics for Comparison.....	Error! Bookmark not defined.
4.2	Structural Comparison of Our Collection with SAP.....	Error! Bookmark not defined.
4.3	Structural Comparison of Our Collection with Other Collection	Error! Bookmark not defined.
5	The Gold Standard Queries.....	Error! Bookmark not defined.
5.1	The Process of Generating Queries.....	Error! Bookmark not defined.
5.1.1	Compute Scores	Error! Bookmark not defined.
5.1.2	Sampling.....	Error! Bookmark not defined.
5.1.3	Manual Annotation:.....	Error! Bookmark not defined.
5.2	The Generated Queries.....	Error! Bookmark not defined.
5.3	The Use of the Gold Standard for BPM Research	Error! Bookmark not defined.
5.4	The Evaluation Metrics.....	Error! Bookmark not defined.
5.4.1	AUC (Area under curve).....	Error! Bookmark not defined.
5.4.2	Precession and Recall.....	Error! Bookmark not defined.
5.4.3	F-measure.....	Error! Bookmark not defined.
6	Conclusion and Future Work	Error! Bookmark not defined.
6.1	Summary of Contribution	Error! Bookmark not defined.
6.2	Fulfilling Objectives	Error! Bookmark not defined.
6.3	Study Findings	Error! Bookmark not defined.
6.4	Limitations	Error! Bookmark not defined.
6.5	Future Work.....	Error! Bookmark not defined.
	References.....	Error! Bookmark not defined.

List of Figures

Figure 1 The steps of data generation briefly.....	Error! Bookmark not defined.
Figure 3: Process model followed proper guideline	Error! Bookmark not defined.
Figure 4: Process model not followed proper guideline.....	Error! Bookmark not defined.
Figure 5: Every split has its join.....	Error! Bookmark not defined.
Figure 6: Mismatch in the process model.....	Error! Bookmark not defined.
Figure 7: Each corresponding gateway type is same	Error! Bookmark not defined.
Figure 8: Each corresponding gateway type is not same.....	Error! Bookmark not defined.
Figure 9: Arcs and gateways not labeled	Error! Bookmark not defined.
Figure 10: Arcs and gateways are labeled	Error! Bookmark not defined.
Figure 11: Each activity has one incoming and one outgoing arc.....	Error! Bookmark not defined.
Figure 12: Activities have more than one incoming and one outgoing arc...	Error! Bookmark not defined.
Figure 13: Loop in a process model	Error! Bookmark not defined.
Figure 14: Unlabeled activities in a process model	Error! Bookmark not defined.
Figure 15: Target less arcs in the process model	Error! Bookmark not defined.
Figure 16: The process models which have created through many pools and lanes. ..	Error! Bookmark not defined.
defined.	
Figure 17: Same examples without any pool or lane.....	Error! Bookmark not defined.
Figure 18: The process with one lane and pool	Error! Bookmark not defined.
Figure 19: The overlapping in a process model	Error! Bookmark not defined.
Figure 20: The claim process model	Error! Bookmark not defined.
Figure 21: The Diameter	Error! Bookmark not defined.
Figure 22: The cut vertices in the above example.	Error! Bookmark not defined.
Figure 23: Original process models.....	Error! Bookmark not defined.
Figure 24: Heavy Revised process model.....	Error! Bookmark not defined.
Figure 25: Steps for Query Generation	Error! Bookmark not defined.
Figure 26: The area under curve	Error! Bookmark not defined.

List of Tables

Table 1: The roles of participants	Error! Bookmark not defined.
Table 2 : Tasks performed by each Researcher	Error! Bookmark not defined.
Table 3: The composition of our process models image collection.....	Error! Bookmark not defined.
Table 4: Comparison between controls flows patterns by wohed et al, 2006	Error! Bookmark not defined.
Table 5: BEST Open Source tool for BPMN 2.0 modeling	Error! Bookmark not defined.
Table 6: Modeling guidelines used in generating data set collection.....	Error! Bookmark not defined.
Table 7: The structural parameters and formulas	Error! Bookmark not defined.
Table 8: Structural metrics related to events and gates.....	Error! Bookmark not defined.
Table 9: The structure of the model	Error! Bookmark not defined.
Table 10: The revisions performed by Researchers.....	Error! Bookmark not defined.
Table 11: The near copy example	Error! Bookmark not defined.
Table 12: Operations performed for generating NC and LR	Error! Bookmark not defined.
Table 13: Number of operations performed to generate NC of 150 BPMN..	Error! Bookmark not defined.
Table 14: The near copy operations performed by R1	Error! Bookmark not defined.
Table 15: The near copy operations performed by R3	Error! Bookmark not defined.
Table 16: The Light Revision example.....	Error! Bookmark not defined.
Table 17: Total Number of operation performed to generate LR of 150 BPMN	Error! Bookmark not defined.
Table 18: The Light Revision operations performed by R2.....	Error! Bookmark not defined.
Table 19: The Light Revision operations performed by R3.....	Error! Bookmark not defined.
Table 20: Total Number of LR operation performed to generate HR of 150 BPMN	Error! Bookmark not defined.
Table 21: LR operations for HR performed by R2	Error! Bookmark not defined.
Table 22: LR operations for HR performed by R3	Error! Bookmark not defined.
Table 23: Operations applied for structural changes	Error! Bookmark not defined.
Table 24: The Total number of structural changes.....	Error! Bookmark not defined.
Table 25: The structural changes performed by R1.....	Error! Bookmark not defined.
Table 26: The structural changes performed by R2.....	Error! Bookmark not defined.
Table 27: The statistics of our collection	Error! Bookmark not defined.

Table 28: The structural comparison results with SAP	Error! Bookmark not defined.
Table 29: Comparison with Heavy Revision parameters with SAP	Error! Bookmark not defined.
Table 30: The minus points	Error! Bookmark not defined.
Table 31: The different model sets	Error! Bookmark not defined.
Table 32: Comparison with Complete Sample	Error! Bookmark not defined.
Table 33: Comparison with Service model	Error! Bookmark not defined.
Table 34: Comparison with Finance model	Error! Bookmark not defined.
Table 35: Comparison with consulting model	Error! Bookmark not defined.
Table 36: structural metrics used for query process	Error! Bookmark not defined.
Table 37 Classification of collection	Error! Bookmark not defined.
Table 38: Classes that have only one process in it	Error! Bookmark not defined.
Table 39: The queries generated from original process models sections	Error! Bookmark not defined.
Table 40: The queries generated from Near Copy process models sections	Error! Bookmark not defined.
Table 41: The queries generated from Light Revision process models sections	Error! Bookmark not defined.
Table 42: The queries generated from Heavy Revision process models section	Error! Bookmark not defined.
Table 43: The list of evaluation metrics	Error! Bookmark not defined.
Table 44: The terms used in Evaluation Metrics	Error! Bookmark not defined.
Table 45: The steps of making Gold Standard	Error! Bookmark not defined.
Table 46: The objectives of our work	Error! Bookmark not defined.
Table 47: Total Queries and Their Results	Error! Bookmark not defined.

