Systematic Literature Review via ML and its representation as Heatmaps

Case in Point: Agile Software Development

Itzik David & Roy Gelbard

Bar-Ilan University - Information System Program

Introduction

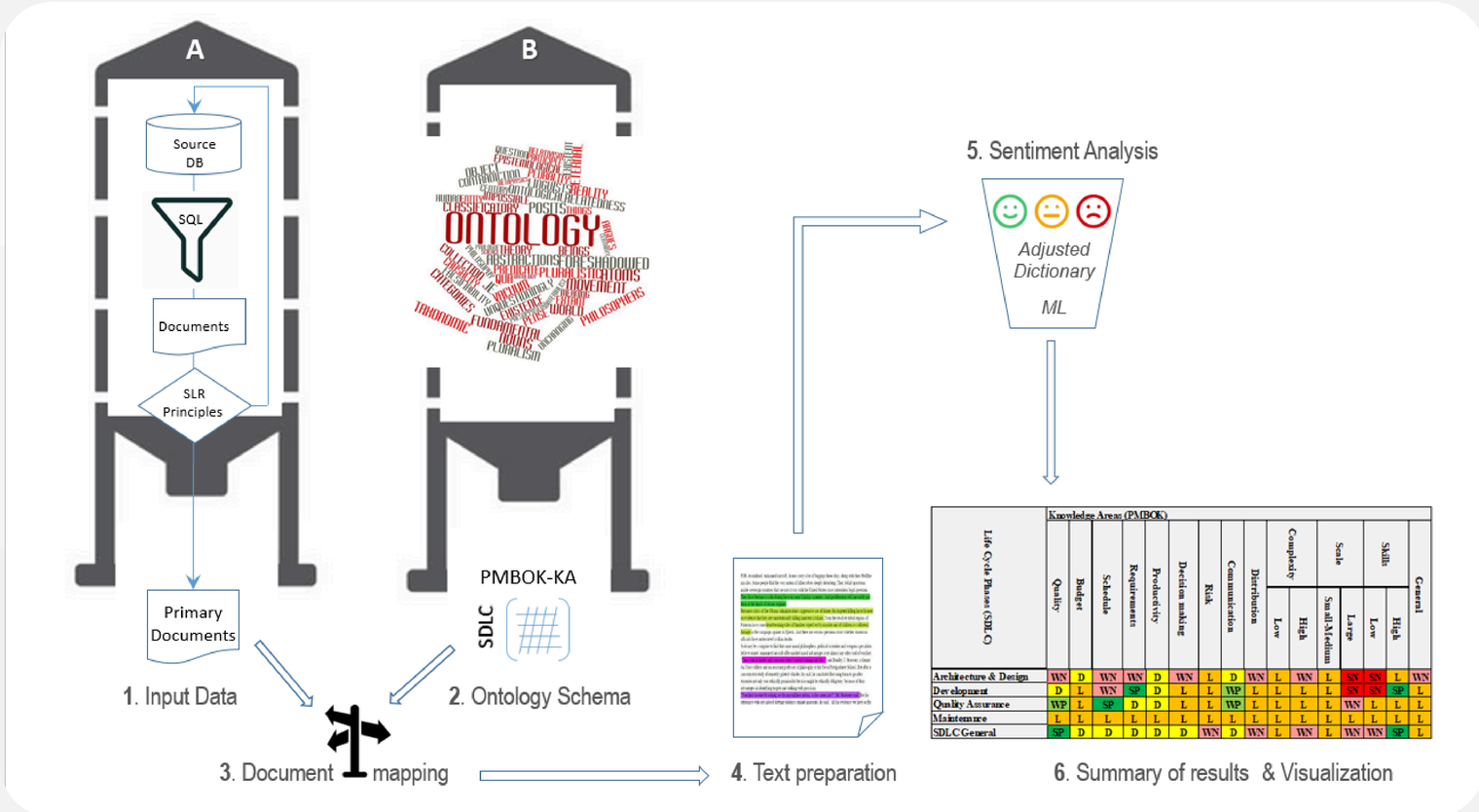
Systematic Literature Review (SLR) is essential for researchers to keep up with past and recent research in their domains. However, knowledge creation has been accelerating in various research domains with a steadily growing number of publications. Researchers, therefore, face the need to examine more publications than in the past. Moreover, most SLRs are performed manually. Staying on top of the state-of-the-art thus becomes a complex and challenging process. The present study introduces a method of utilizing Machine Learning (ML) techniques such as text mining, sentiment analysis, and visualizations to make SLR more effective. The results reveal an impressive accuracy in sentiment analysis evaluation within the agile software development domain. They also illustrate the method's technical advantages and efficiency in utilizing resources, time, and standard evaluation that prevents biases.

Visual Abstract

Silo A = Fundamental stages of the manual research on which the current research is based. The documents identified in the SLR process as primary documents are represented as #1.

Silo B = Future research that is planned to focus on the possibility of automating ontology identification. The ontology defined in the manual SLR research consists of the knowledge areas defined in PMBOK and SDLC stages. The combination of these two components creates the ontology schema matrix (#2).

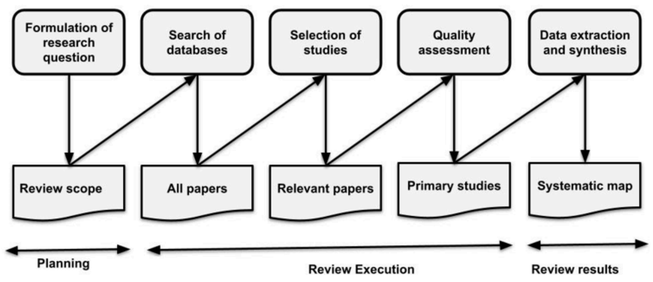
1 = Primary documents, which serve as input data for current research.
2 = Ontology adopted based on the manual SLR research.
3 = Mapping documents to the relevant cells in the ontology matrix.
4 = Identification of position sentences in each document.
5 = Sentiment analysis of each position sentence.
6 = Visualization of the findings within the ontology matrix

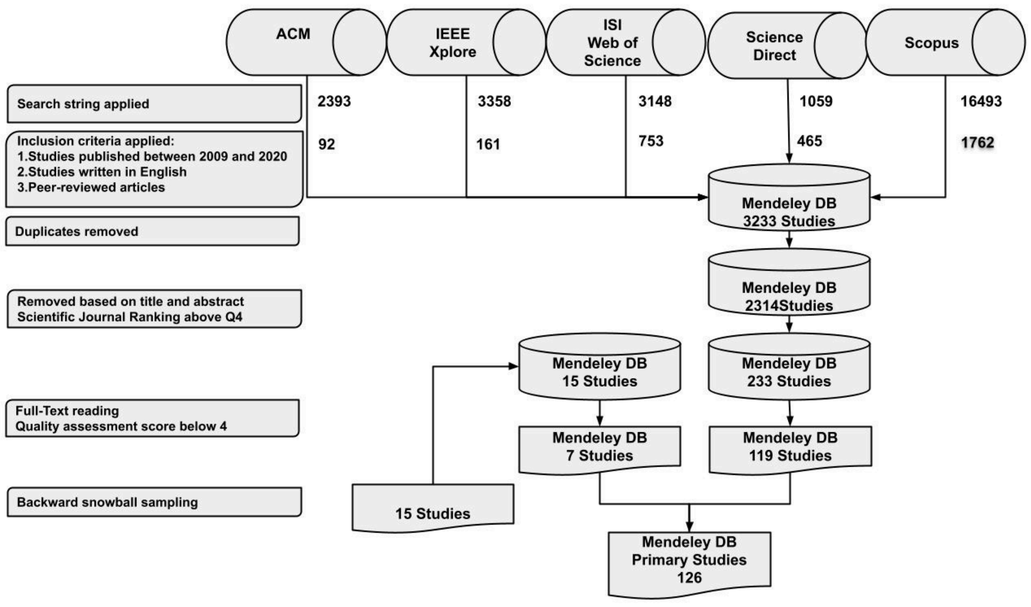


Methodology

The research unfolds in three pivotal stages:

1. Sentiment analysis is manually performed through a Systematic Literature Review process within the agile software development domain. This process creates a labeled dataset, laying the groundwork for the subsequent phase - an experiment.
2. Sentiment analysis is automated through machine learning techniques, with dictionaries and algorithms tailored to the context of agile software development. The accuracy of the machine-generated sentiment evaluated by conventional metrics and then then compared to the manually analyzed sentiments to validate its reliability.



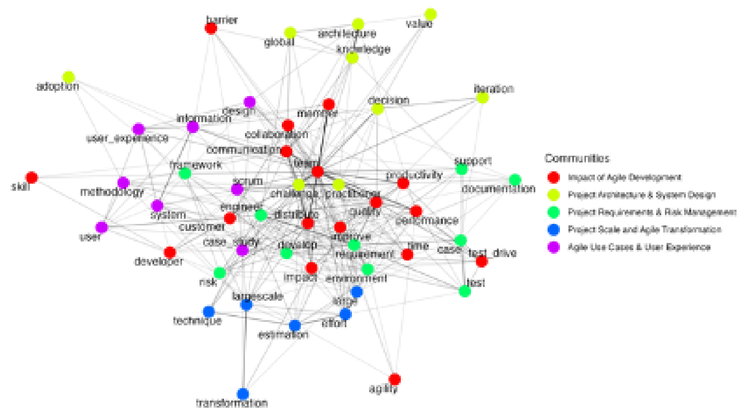


Topic Modeling Analysis - Latent Dirichlet Allocation & Community Detection

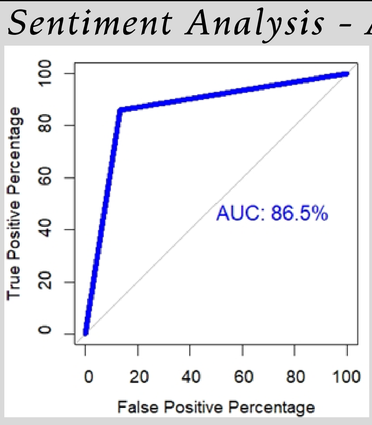
Topic 1	Topic 2	Topic 3
Collaboration & Communication	*Software Engineering Processes & Methodologies*	*Quality & Impact*
team	team	challenge
quality	methodology	requirement
communication	decision	practitioner
productivity	customer	communication
test_drive	engineer	team
impact	user_experience	quality
develop	risk	architecture
performance	estimation	knowledge
environment	value	design
framework	effort	distribute
largescale	requirement	case_study
transformation	scrum	system
increase	distribute	interview
case_study	user	engineer
challenge	develop	case

Cross Validation

Variables supported/not supported by both topic modeling methods



Sentiment Analysis - AUC & Confusion matrix of sentiment classification



Actual \ Predicted	Positive	Negative	Total	Recall
Positive	457	58	515	88.74%
Negative	69	357	426	83.80%
Total	526	415	941	86.50%
Precision	86.88%	86.02%	86.50%	-

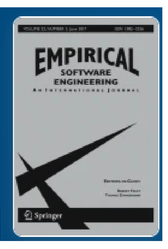
Life Cycle Phases (SDLC)	Quality	Budget	Schedule	Requirements	Productivity	Decision making	Risk	Communication	Distribution	Complexity: Low	Complexity: High	Scale: Small-Medium	Scale: Large	Skills: Low	Skills: High	General
Architecture & Design	0	0	0	0.06	0.25	0	L	0.15	0.25	L	0.3	L	0.07	0.1	L	0.4
Development	0.31	L	0.2	0	0	L	L	0.26	L	L	L	L	0	0.29	0.38	L
Quality Assurance	0.06	L	1.60	0.29	0.06	L	L	0	L	L	L	L	0	L	L	L
Maintenance	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
SDLC General	0	0.17	0.05	0	0	0.25	0.29	0.04	0.12	L	0	L	0.2	0.07	0.19	L

Impact on findings: None, Low, Medium, Lack of evidence, High, Very High

Published: 14 July 2023 Volume 28, article number 105, (2023)

Does agile methodology fit all characteristics of software projects? Review and analysis

David Itzik & Gelbard Roy



WIREs DATA MINING AND KNOWLEDGE DISCOVERY

Using Machine Learning for Systematic Literature Review Case in Point: Agile Software Development

Itzik David, Roy Gelbard

Vol 15, Issue 1
March 2025