Slovak University of Technology in Bratislava

Faculty of Informatics and Information Technologies

Ing. Viktor Matovič

Dissertation Thesis Abstract

Establishing Pattern Sequences Using Probabilistic Methods

to obtain the academic title of *Philosophiae doctor* (PhD.)

Study Program:	Applied Informatics
Field of Study:	Computer Science
Place of development:	Institute of Informatics, Information Systems and Software Engineering Faculty of Informatics and Information Technologies Slovak University of Technology in Bratislava

Bratislava, May 2025

Dissertation Thesis has been developed at the Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava.

Submitter: Ing. Viktor Matovič
Institute of Informatics, Information Systems
and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava

Supervisor: Doc. Ing. Ján Lang PhD.
Institute of Informatics, Information Systems
and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava

Former Supervisor and Consultant: Prof. Ing. Valentino Vranić PhD.
Faculty of Informatics
Pan-European University, Bratislava, Slovakia

Oponents: doc. Ing. Miroslav Bureš Ph.D.
Faculty of Electrical Engineering
Czech technical university in Prague, Czech Republic

MSc. Marko Ranković PhD.
Faculty of Electrical Engineering and Informatics
Union Nikola Tesla University, Republic of Serbia

Dissertation Thesis Abstract was sent: ....
Dissertation Thesis Defence will be held on .... at .... pm at the Institute of Computer Engineering and Applied Informatics, Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava (Ilkovicova 2, Bratislava).

Prof. Ing. Ivan Kotuliak, PhD.
Dean of FIIT STU in Bratislava

ANNOTATION

Slovak University of Technology Bratislava
Faculty of Informatics and Information Technologies

Author:	Ing. Viktor Matovič
Supervisor:	Doc. Ing. Ján Lang PhD.
Former Supervisor and Consultant:	Prof. Ing. Valentino Vranić PhD.
Degree Course:	Applied informatics
Title:	Establishing Pattern Sequences Using Probabilistic Methods
May 2025	

Organizational patterns can be applied to correct problems or build a new organization from scratch. Security patterns can help organizations develop secure software. Security in the organization can be established using sequences of organizational and security patterns. Although isolated applications of one or more security patterns are typical in practice, an effective response to security threats requires the utilization of an entire pattern catalog. Composing software patterns in a particular sequence order makes them more efficient. Solutions for establishing sequences of security and organizational patterns are often documented in plain text using explicit pattern relationships. Identifying the most probably used security and organizational patterns from their text descriptions allows us to establish meaningful sequences from these patterns. We show that the strongest symmetries between organizational and security patterns can be used to establish meaningful pattern sequences that are expected to be used and are established without using statistics about the past use of these patterns in practice. We show how the strongest symmetries of relationships between patterns can be extracted using stochastic processes to establish pattern sequences that stay meaningful even if the order of their subsequently used patterns is reversed. We show how these pattern sequences can solve security problems in organizations. We show that symmetries of relationships that can be used to establish meaningful pattern sequences can be identified from text descriptions of any patterns.

ANOTÁCIA

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

Autor:	Ing. Viktor Matovič
Vedúci Práce:	Doc. Ing. Ján Lang PhD.
Predchádzajúci Vedúci Práce a Konzultant:	Prof. Ing. Valentino Vranić PhD.
Študijný program:	Aplikovaná Informatika
Názov Práce:	Zakladanie Sekvencií Vzorov Použitím Pravdepodobnostných Metód
May 2025	

Organizačné vzory možno použiť na vyriešenie problémov alebo na vybudovanie novej organizácie. Bezpečnostné vzory pomáhajú organizáciám pri vývoji bezpečného softvéru. Bezpečnosť v organizácii možno zabezpečiť pomocou sekvencií organizačných a bezpečnostných vzorov. Hoci izolované použitie jedného alebo viacerých bezpečnostných vzorov je v praxi typické, efektívna reakcia na bezpečnostné hrozby si vyžaduje použitie celého katalógu vzorov. Kombinácia softvérových vzorov v určitom poradí v sekvenciách ich robí efektívnejšími. Riešenia na zakladanie sekvencií vzorov sú často zdokumentované v textových opisoch vzorov pomocou ich explicitných vzťahov. Identifikácia najpravdepodobnejšie použitých vzorov z ich textových popisov umožňuje vytvoriť z týchto vzorov zmysluplné sekvencie. Táto dizertačná práca ukazuje, že najsilnejšie symetrie medzi organizačnými a bezpečnostnými vzormi možno použiť na vytvorenie zmysluplných sekvencií vzorov ktoré sú očakávané byť použité a to bez použitia štatistík o minulom použití týchto vzorov v praxi. Táto dizertačná práca ukazuje spôsob ako možno najsilnejšie symetrie vzťahov medzi vzormi extrahovať pomocou stochastických procesov na založenie sekvencií vzorov, ktoré zostávajú zmysluplné, aj keď sa poradie ich následne použitých vzorov obráti. Práca ukazuje ako je možné použiť symetrie vzťahov na založenie zmysluplných sekvencií vzorov a ako je možné identifikovať tieto symetrie vzťahov z textových popisov akýchkoľvek vzorov. Práca predstavuje sekvencie kombinujúce organizačné a bezpečnostné vzory ktoré riešeia vybrané bezpečnostné problémy v organizáciách.

Table of Contents

1. Introduction

Patterns are an organized way of exchanging best practices between actors in a specific domain. Software developers dealing with commonly recurring problems converge to use similar practices, documented and reused as patterns. A specific category of patterns is organizational patterns, which can be considered best practices for organizing people working together, building companies, or writing program code. Another specific category of patterns is security patterns, which can be used to implement security in software systems.

Patterns are often documented in pattern languages or catalogs, but these sources of patterns do not identify combinations of patterns that are expected to be used together by pattern users. Text descriptions of software patterns are also often published in isolated sources, without information on how to combine them to challenge complex problems related to certain application domains.

Not all pattern relationships are explicitly stated in the text, and some remain hidden, which are called implicit relationships. Patterns may be combined using implicit relationships. Implicit relationships can be extracted from pattern descriptions. Many implicit relationships between patterns exist. Some implicit relationships can be used to identify patterns that can be combined in meaningful pattern sequences.

Multiple patterns can be combined in pattern sequences. Pattern sequences help solve complex problems that cannot be solved using a single pattern. Establishing pattern sequences helps solve specific problems effectively, even if the problem is complex, rather than using patterns individually. To solve more significant problems, there is a need to know how to establish pattern sequences. Probabilities with which implicit relationships between patterns would establish meaningful pattern sequences can be used to distinguish pattern sequences that are expected to be used in practice.

According to Sousa et al. [11], verbs used to express relationships between patterns do not imply (joint or conditional) correlation or causality. Waseeb et al. [12] note that not all pattern sequences provide meaningful ways to solve a problem or build a system entirely from scratch Meaningful pattern sequences based on conditional dependencies between patterns cannot be established by just reading their text descriptions. One cannot be sure about using any pattern at any given time. By calculating the probabilities with which patterns would be applied together, these probabilities can be used to establish meaningful pattern sequences. Conditional dependencies between patterns can be extracted from their descriptions. These dependencies can be used to identify patterns that are expected to be combined in pattern sequences. Conditional dependencies between patterns can be identified by calculating the conditional probabilities of the combined use of patterns. Although some conditional dependencies between patterns can be identified by reading text descriptions of these patterns, this is hard to do when conditional dependencies must be identified from an extensive number of patterns. This way, a lot of conditional dependencies would remain hidden and would not be used to identify solutions presented by viable pattern sequences. Conditional probabilities of the combined use of patterns can be calculated efficiently using stochastic processes. Stochastic processes can also be used to discover previously unseen pattern sequences that provide innovative solutions to certain problems. According to Gareth et al. [8], decisions that use patterns may vary and can be explained with probability-based models. Conditional dependencies between patterns can also be used to construct a probability model, such as a Bayesian network [9]. This model can show which pattern sequences are expected to be used and which are not.

The thesis presents three methods to establish pattern sequences that are expected to be used and are established using implicit relationships extracted from text descriptions of patterns. Pattern sequences can be characterized as expected to be used if the probability of their use is higher than the probability of use of other pattern sequences. Characterizing pattern sequences as expected enables pattern users to distinguish their inherent properties and the probability of their use.

The presented thesis aims to understand the sequential nature of the patterns by establishing sequences of patterns that are expected to be used. It is assumed these sequences would help identify the underlying forces behind patterns and how they shift. These forces can be compared to real organizations and may give insight into the human dynamics that help or hinder the effectiveness of teams.

1.1 Thesis Statement

Pattern sequences combine patterns to solve complex problems that cannot be solved using single patterns. Using only explicit relationships between patterns to establish pattern sequences prevents pattern users from establishing some sequences that might be discovered as expected to be used. There must be a way to establish pattern sequences from any patterns, because expected solutions to complex problems can be needed in many application domains. Some of these pattern sequences are more likely to be used than others; therefore, any method establishing pattern sequences must be able to distinguish which are most expected to be used, which are expected to be used less, and which are unexpected and should be avoided. Currently, no methods of establishing pattern sequences that are expected to be used in practice exist. Currently, no methods of establishing pattern sequences that are expected to be used and that would not establish pattern sequences using software patterns from one particular application domain selected by the authors of these methods exist. Methods that would not require changing text descriptions of these patterns, which would not use statistics about the past use of patterns in practice, and which would use conditional dependencies between these patterns to establish pattern sequences, do not exist. Therefore, the thesis statement is as follows:

> Relationships between software patterns that lead to the identification of meaningful pattern sequences can be discovered from their descriptions via probabilistic methods. Some of these relationships lead to identifying pattern sequences expected to be used in practice.

1.2 Contributions

The presented thesis brings the following contributions that support the thesis statement presented in the previous section:

- **Method of establishing pattern sequences using stochastic trees (Section 4.0.1)** can be used to establish pattern sequences without information about the past use of patterns, that are expected to be used, and that can be combined to solve complex problems.

- **Method of establishing pattern sequences using Bayesian networks (Section 4.0.2)** can be used to establish sequences of conditionally dependent patterns that are expected to be used and that can be combined to solve complex problems.

This method can be used to calculate the likelihood of use of established pattern sequences.

- **Method of establishing pattern sequences using artificial neural networks (Chapter 5)** automates establishing expected pattern sequences, which are expected to be used, without requiring text descriptions of these patterns to link to each other.

None of these methods needs to change text descriptions of patterns before establishing pattern sequences. All methods use conditional probabilities of using patterns to establish pattern sequences. The method of establishing pattern sequences using Bayesian networks solves design problems of the Method of establishing pattern sequences using stochastic trees. The method of establishing pattern sequences using artificial neural networks automates the calculation of the probabilities of using patterns to establish pattern sequences.

2. How Patterns Are (Not) Connected

Patterns are an organized way to exchange best practices between actors in a specific domain area [1]. Software developers who deal with commonly recurring problems converge on using similar solutions described as practices, documented, and exchanged as patterns.

Software patterns are documented either in isolation, without providing links to other related patterns, or by providing many links to related patterns without information on which links to follow.

Complex problems can be solved by combining these patterns into pattern sequences. Patterns can be combined into sequences using links extracted from their text descriptions. Following some explicit links between patterns does not establish meaningful pattern sequences. By calculating probabilities that explicit links in descriptions of software patterns would be followed, combinations of patterns that are likely to be used can be identified. Patterns combined in pattern sequences that are expected to be used could uniquely solve complex problems, even crossing multiple domains. Methods to establish pattern sequences that are expected to be used are rarely provided by authors documenting patterns.

Apart from pattern languages, patterns are also organized in pattern catalogs, pattern collections, and web repositories [3]. These sources can be used to extract explicit links between software patterns. There are multiple links between software patterns without any information about which are expected to be followed. Probabilities of using related patterns together can be calculated, and these probabilities can be used to establish meaningful pattern sequences.

Coplien and Harrison [5] documented four languages of organizational patterns that can be used for the growth of companies or to repair their internal structure or processes. Meaningful pattern sequences can be established from organizational patterns. No method to establish sequences of organizational patterns is provided by Coplien and Harrison. Coplien and Harrison do not provide a hint on how to establish pattern sequences using patterns from more than two pattern languages.

Cordeiro et al. [6] documented a catalog of 106 security patterns that can be used to enhance the security of information systems or implement security guidelines. Meaningful pattern sequences can be established from security patterns to implement security in an organization or security guidelines in software systems. No security pattern in their catalog is marked as the one that must be applied. There are always numerous related security patterns that can be applied under different conditions.

2.1 Pattern Sequences Expected To Be Used

To solve more significant problems with sequences of organizational patterns, it must be known which patterns to use and compose together. Nobody can be sure about using any particular pattern at any given time. This is why there is a need to calculate the probability to decide which patterns are expected to be applied together.

The probability of application of pattern by the assumed domain expert can be calculated using the frequentist, Bayesian approach presented in the work of Kruschke [10], Barber [2], Koller and Friedman [9] or through sequence models such as Markov chain presented by Zhang et al. [13].

The probability of application of the pattern can be calculated using the frequentist approach as $1/N$, where the denominator N is the number of patterns documented in the pattern repository. This means the probability of using each pattern is the same, and this probability cannot be used to select the next pattern with the highest probability to be used. The suitability of various stochastic processes for the calculation of probabilities of applying patterns must be calculated.

Probabilities of use of each applicable pattern need to be calculated to determine which pattern has the highest probability of being applied. The probability of the use of the pattern sequence established based on the selection of the pattern with the highest probability of use needs to be calculated to determine if the resulting pattern sequence is expected to be applied.

2.2 Establishing Pattern Sequences Using Stochastic Processes

The probability of applying a pattern in a pattern sequence can be calculated with a frequentist statistical approach. If it is possible to observe applications of N patterns many times, the frequency of application of each will be near $(1/N) * 100$ %. This means the probability of applying each pattern in the pattern language or catalog would be the same. This probability must be calculated using sophisticated and complex mathematical models. Patterns from languages or catalogs with many patterns are less likely to be used because the denominator N in $1/N$ is very high. If probabilities of pattern use are *a priori* the same, this probability cannot be used to establish the pattern sequences that are expected to be used, because it is impossible to identify a pattern with the highest probability of use.

Mass use of patterns is problematic to observe. Because of this, the probability of use of the pattern can be calculated using a Bayesian approach, an analytical approach by employing the Beta distribution presented in another use case by Kruschke [10], or Markov Chain Monte Carlo method implemented with the Metropolis algorithm presented using another example in the work of Kruschke [10].

2.2.1 Bayesian Approach

The Bayes rule, often used in the Bayesian statistical approach, can be used to calculate the probability of applying a pattern, considering contradicting forces that this pattern tries to resolve or previous applications of patterns. The Bayesian approach can be used to calculate the probability of use of a pattern or a pattern sequence, even though the real use of patterns is impossible to observe. If the Bayesian approach to calculating the probability of pattern use is used, it means it is impossible to identify the pattern most expected to be used next because a pattern trying to resolve similar contradicting forces is expected to be used with the same probability. This means that the Bayesian approach should be combined with another stochastic process to identify the pattern that is expected to be used.

2.2.2 Beta Distribution

The probability of applying a pattern can be calculated using the Beta distribution function. Information about how frequently the patterns are used in the software industry can be provided as input to the Beta distribution function. The frequency of pattern use can be calculated from pattern stories, like those provided by Buschmann et al. [3] or Kruschke [10]. The pattern that is most expected to be used can be identified by calculating the probabilities of using patterns that are candidates for expected ones. These probabilities can be subsequently compared to select the most probable option. The problem with this approach is that calculating the probability of applying patterns results in small numbers, even for a small number of applicable patterns. This may obscure the true usefulness of patterns only because a small probability was calculated for them.

2.2.3 Metropolis Algorithm

Because the number of previous pattern applications of patterns can be extracted from pattern stories, the problem of having small numbers can be solved by applying the Metropolis algorithm. It can be used to transform the prior probability into a posterior probability that any pattern from a set of N patterns would be applied, considering the previous application of this pattern Z times. The problem with using this algorithm is that it can only be used to establish pattern sequences expected to be used if there is information about the past use of patterns.

Hazen's [7] stochastic trees can be used to calculate the probability of application of patterns in sequences. The stochastic tree can also be used to calculate probabilities of using sequences of patterns. Probabilities of use of pattern sequences calculated using the stochastic tree can be compared, and the sequence with the highest probability can be selected as the most expected to be used. The disadvantage of using these trees to find the most probably used pattern sequences is that there may be more than one pattern sequence with the highest probability of use. The problem with using stochastic trees to calculate the probability of use of patterns or pattern sequences is that it does not consider conditional dependencies and conditional independencies between patterns. Conditional dependencies and independencies between patterns can be identified using Bayesian networks [2]. Stochastic trees should be used with another stochastic process to establish pattern sequences that are expected to be used.

2.2.4 Grid Approximation

The probability of applying patterns or pattern sequences can be calculated with the grid approximation technique presented in the work of Kruschke [10]. This technique can be used to establish pattern sequences by selecting the next applicable pattern that has the highest probability of use from all linked patterns in the text description. The table in Figure 2.1 shows an example of this calculation on the seven organizational patterns documented by Coplien and Harrison [5]. The organizational patterns in this table are those mentioned in the Skunk Works text description provided by Coplien and Harrison [5] (all are applicable after Skunk Works).

The table in Figure 2.1 shows absolute differences between the conditional probability of use of the pattern in a row after the pattern in the column, and the opposite probability to it calculated with the Bayes rule. Conditional probabilities that served as input for these calculations were extracted from the stochastic tree constructed for the meaningful pattern sequence consisting of patterns in the Piecemeal Growth pattern language from Coplien and Harrison [5], which is also available in the GitHub repository. [1].

The probability of using any pattern can be calculated by summing up the row or column in the table in Figure 2.1. Calculating the probability of the use of the pattern in Figure 2.1 by marginalizing the row or column is a more accurate indicator of the credibility

[1] stochastic tree is available at `https://github.com/viktorFIIT/fiit-research-resources/blob/main/stochastic-tree/stochastic-tree-piecemeal-growth-pattern-language.pdf`

of the pattern because it outputs the probability of the use of the pattern irrespective of other patterns that this pattern refers to.

The probability of subsequent use of the patterns in the row and column can be calculated by dividing the value of a cell in the table in Figure 2.1 by the marginal probability for a given row or column (depending on the pattern for which the probability is calculated).

Pattern	Self Selecting Team	Compensate Success	Gate Keeper	Patron Role	Marginal Probability
Skunk Works	0.00784	0.00596	0.08227	0.45103	0.00784+0.00596+0.08227+0.45103=0.5471
Size The Organization	0.0058	0.2878	0.0729	0.00596	0.0058+0.2878+0.0729+0.00596=0.37246
Fire Walls	0.0622	0.2878	0.0729	0.0622	0.0622+0.2878+0.0729+0.0622=0.4851
Marginal Probability	0.00784+0.0058+0.0622=0.07584	0.00596+0.2878+0.2878=0.58156	0.08227+0.0729+0.0729=0.22807	0.45103+0.00596+0.0622=0.51919	

Figure 2.1: Differences Between Conditional Probabilities of Use of Organizational Patterns

Pattern sequences most expected to be used together can be established with the combination of the Hazen's [7] stochastic trees and Grid approximation. The problem with using the Grid approximation is that it does not consider conditional dependencies and independencies between software patterns. Because of this, grid approximation should be combined with another technique to establish pattern sequences expected to be applied.

2.2.5 Artificial Neural Networks

The softmax regression model can be used to compute the probability that a vector of words can be ascribed to the name of the pattern. This implies that the softmax regression model can be used to compute the probability that the sentence describing any problem can be ascribed to the name of the organizational pattern. The softmax regression model [13] designed to compute these probabilities, can be tested on text descriptions of patterns from Coplien and Harrison [5], if it can link parts of text descriptions to the names of these organizational patterns. Probabilities with which parts of text descriptions of patterns or descriptions of problems can be linked to names of patterns can be sorted in descending order, and this order can be used to establish pattern sequences.

The probability with which the part of the description of a problem can be linked to the name of the pattern is a form of an implicit relationship between patterns.

3. Implicit Relationships Between Patterns

Sousa et al. [11] identified patterns expected to be used together by calculating symmetric relationships between two software patterns as differences between conditional probabilities of their subsequent use. Symmetries are another implicit relationship between patterns that can be used to establish pattern sequences. Sousa et al. [11] did not identify conditional dependence and independence between patterns that could otherwise be used to find the most expected order of use of patterns in pattern sequences. Bayesian belief networks from Barber [2] can be used to identify conditional dependence between more than two patterns to establish meaningful pattern sequences. In contrast to the work of Sousa et al. [11], domain experts do not have to provide conditional probabilities to calculate symmetries of relationships between patterns. Symmetries of relationships between patterns can be calculated using probabilities extracted from Hazen's [7] stochastic trees constructed on top of any pattern sequences. The strongest symmetries of relationships can also be found between security patterns from the pattern catalog of Cordeiro [6] or organizational patterns documented by Coplien and Harrison [5]. Symmetries of relationships between these patterns can be used to establish meaningful pattern sequences that combine patterns from multiple domains.

None of the analyzed methods and techniques using implicit relationships to identify commonly used patterns support calculating probabilities of using pattern sequences and comparing these probabilities to identify pattern sequences that are expected to be used. None of these research works presents methods to establish pattern sequences with a high probability of their use. None of the research works presents methods to establish sequences of patterns from any domain.

In this thesis, it was implied that the sequences of software patterns need to be established:

- Considering the strength of symmetry of relationships that was shown to identify patterns that are expected to be used together.
- Considering conditional dependence between patterns because conditionally dependent patterns are expected to be used together.
- Without the need to update text descriptions of patterns.
- Using any software pattern allowing the establishment of pattern sequences that combine patterns from multiple domains and thus provide unique complex solutions to complex problems.
- Without statistics about the past use of software patterns, because these may not be available.

4. Establishing Pattern Sequences Using Stochastic Processes

Pattern sequences that are expected to be used can be established using stochastic processes. Pattern sequences are expected to be used because they are established from patterns that have the strongest symmetry of relationship between each of their two subsequently used patterns. This symmetry is calculated using conditional probabilities of using conditionally dependent patterns. This PhD thesis presents the following four experiments with combinations of stochastic processes to establish pattern sequences.

Assuming that symmetries of relationships between patterns can be used to establish meaningful pattern sequences, multiple experiments were conducted to identify the best order in which symmetries of relationships between patterns and the two following stochastic processes would be used to establish pattern sequences:

- Hazen's stochastic trees [7] - because they can be used to identify pattern sequences with the highest probability of their use.
- Bayesian networks [2] - because they can be used to identify conditionally dependent and independent patterns. Conditionally dependent patterns are expected to be used in pattern sequences.

4.0.1 The Method of Establishing Pattern Sequences Using Stochastic Trees

A method of establishing pattern sequences using stochastic trees was designed here and executed in two experiments. In the first experiment, the method was applied to organizational patterns from Coplien and Harrison. [5] In the second experiment, the method was applied to security patterns from the pattern catalog of Cordeiro et al. [6]

In both experiments, Hazen's stochastic trees were constructed on top of pattern sequences established using explicit relationships between organizational patterns from Coplien and Harrison [5] and security patterns from the pattern catalog of Cordeiro et al. The pattern sequence with the highest probability was identified from the stochastic tree. This pattern sequence was the candidate for the pattern sequence that is expected to be used. Every two organizational or security patterns in this candidate were checked to see if they have the strongest symmetries of relationships. The strength of symmetry of relationship $|p(related\,pattern|candidate\,pattern) - p(candidate\,pattern|related\,pattern)|$ was calculated between patterns in the candidate and their related patterns found in their text descriptions. Strengths of symmetries of relationships between patterns were calculated using probabilities of subsequent use of patterns extracted from stochastic trees.

Pattern sequences that are most expected to be used and meaningful were established following the strongest symmetries of relationships between patterns. Pattern sequences that are unexpected to be used were established following the lowest strength of symmetry of relationships between patterns. A few unexpected pattern sequences were found meaningful. Unexpected pattern sequences represented unfinished solutions to certain problems and required more patterns to be applied.

Evaluation

All pattern sequences most expected and expected to be used and established using the method of establishing pattern sequences using stochastic trees were found to be used in practice. It was also found that all pattern sequences that are most expected and expected to be used could be described using a pattern story (anecdote). Evaluating pattern sequences by describing them in pattern stories is a common evaluation technique for patterns and pattern sequences.

The uniqueness of one of the pattern sequences most expected to be used was evaluated to check that it can be established in only one way. Alternative patterns for those used in the most expected to be used pattern sequence Skunk Works → Self Selecting Team → Diverse Groups solving similar contradicting forces were searched in the work of Coplien and Harrison [5]. If they were found, alternative pattern sequences to the most expected sequence were established. A stochastic tree with a chain rule was used to calculate the probability of applying the pattern sequence that is most expected to be used and the alternative pattern sequence to it. It was found that Build Prototypes is more likely to be used than Skunk Works in this pattern sequence (but both solve similar contradicting forces).

Pattern sequences most expected and expected to be used were evaluated to determine if the symmetric relationships between the patterns they consist of form symmetry groups in a similar way to Coplien and Zhao [4] evaluated symmetry of design patterns. It was found that all symmetric relationships used to establish these pattern sequences, which were found to be most expected and expected to be used, form symmetry groups.

The meaningfulness of the five conditional probabilities of pattern use used to calculate the strength of symmetry of the relationship was evaluated by comparing them to the probabilities of pattern use in the opposite order and by checking if the differences between these two probabilities are meaningful. It was found that differences between the five conditional probabilities used in the calculations of strengths of symmetry of relationships between organizational patterns in most expected to be used and expected to be used pattern sequences represent meaningful relationships.

A domain expert checked the strengths of symmetry of relationships between security patterns in a pattern sequence that is most expected to be used to see if they represent meaningful relationships. Evaluation of the meaningfulness of all four relationships in this sequence was successful.

Results

One pattern sequence that is most expected to be used, four pattern sequences that are expected to be used, and two pattern sequences that are unexpected to be used were established from organizational patterns. Two pattern sequences that are most expected to be used, three pattern sequences that are expected to be used, and three pattern sequences unexpected to be used were established from security patterns.

The strongest symmetries of the relationships establish pattern sequences that are most expected to be used and that stay meaningful even if the order of their subsequently used patterns is reversed.

One sequence of organizational patterns that is most expected to be used could be combined with one sequence of security patterns that is most expected to be used to establish security in the organization. One sequence of organizational patterns that is expected to be used could be combined with one sequence of security patterns that is also expected to be used to establish security in the organization.

4.0.2 The Method of Establishing Pattern Sequences Using Bayesian Networks

A method of establishing pattern sequences using Bayesian networks was designed here and executed in three experiments. In the first experiment, the method was applied to organizational patterns from Coplien and Harrison. [5] In the second and third experiments, the method was applied to security patterns from the pattern catalog of Cordeiro et al. [6]

In a first experiment, a meaningful pattern sequence was established using all patterns in the Piecemeal Growth pattern language from Coplien and Harrison [5] to learn about the piecemeal growth of an organization. It was established using explicit relationships between its patterns to establish all pattern sequences that are most expected to be used. Hazen's stochastic tree [7] was constructed to calculate probabilities of use of patterns in this meaningful pattern sequence. The selected meaningful pattern sequence was checked to see if another pattern could be used between its two subsequent patterns. If the pattern to be included was found, the strength of symmetry of the relationship between this pattern and the pattern used before and after it in the meaningful pattern sequence was quantified using probabilities extracted from the stochastic tree. If the symmetry of the relationship was weaker than the strength of symmetry of the relationship between the pattern used before and after the included pattern, the included pattern was not expected to be used between these two patterns. If the symmetry of the relationship was stronger than the strength of symmetry of the relationship between patterns used before and after the included pattern, the pattern sequence of three patterns was established.

In the second and third experiments, Hazen's stochastic trees [7] were constructed to calculate probabilities of initial pattern sequences established using explicit relationships between security patterns. Initial pattern sequences established using explicit relationships between security patterns were modeled in a Bayesian network. It was found inside this pattern sequence that each two subsequently used patterns are conditionally dependent and have the strongest symmetry of relationship.

Evaluation

Evaluation concluded that the two most expected to be used sequences of organizational patterns established in the first experiment can be successfully described with pattern stories. One of the sequences of security patterns established during the second experiment, which was found to be the most expected to be used, was successfully described with a pattern story.

Evaluation concluded that transitive dependencies between security patterns exist in one of the three pattern sequences found to be most expected to be used and established in the second experiment. It was found that there is only one way that one of the three pattern sequences found to be most expected to be used and established during the second experiment would be established. Evaluation concluded that implicit relationships exist between two subsequent patterns in one of the three pattern sequences most expected and established in the second experiment.

By calculating the probability of the use of the five pattern sequences found to be most expected and established in the third experiment, it was found that it is more probable that the order of their patterns is not likely to change. The probabilities of using pattern sequences calculated during the third experiment with the stochastic trees and Bayesian belief networks were also checked to see that there are no other probabilities that would imply these pattern sequences are not the most expected to be used.

The existence of transitive dependencies was confirmed between one sequence of security patterns, found to be most expected to be used.

Results

The Bayesian network was employed to calculate the likelihood of the implicit relationships between organizational patterns, given that these relationships serve as preconditions for applications of subsequent patterns in these sequences.

During the first experiment, the two pattern sequences most expected to be used in practice were established from organizational patterns. Two sequences of organizational patterns that are unexpected to be used and shall be avoided were also established.

During the second and third experiments, the Bayesian belief network was used to identify conditionally dependent and independent security patterns. This conditional dependence was used to identify security patterns expected to have the strongest symmetry of relationships between. Conditionally independent patterns were expected to be used individually or avoided in the pattern sequences unexpected to be used.

Three pattern sequences with security patterns were established during the second experiment. These pattern sequences are most expected to be used because they have a higher probability of use than any other sequences, and any two subsequently used patterns had the strongest symmetry of relationship. Eight additional pattern sequences are not meaningful and are unexpected to be used because they have the weakest symmetry of relationship between any two subsequently used patterns.

Five sequences of security patterns that are most expected to be used were established in the third experiment. These pattern sequences have a higher probability of use than any other sequences, and any two subsequently used patterns in them had the strongest symmetry of relationship. Seven additional sequences of security patterns were established by graphically examining the visualization of a Bayesian network modeling explicit relationships between security patterns and identifying conditionally dependent patterns. Two sequences of security patterns that are unexpected to be used were established by identifying conditionally independent patterns from the visualization of a Bayesian belief network.

One sequence of organizational patterns found to be most expected to be used and established during the first experiment could be combined with the two sequences of security patterns found to be most expected to be used and established during the second and third experiments. One additional sequence of organizational patterns found to be most expected to be used and established during the first experiment could be combined with one sequence of security patterns established during the third experiment. The resulting combined sequences could be used to establish security in the organization.

5. Establishing Pattern Sequences Using Artificial Neural Networks

A method of establishing pattern sequences using artificial neural networks was designed here and executed in four experiments with organizational patterns from Coplien and Harrison [5].

Artificial neural networks were used to calculate the probabilities that sequences of terms describe software patterns. These probabilities were used in selecting patterns that can be used to solve complex problems in practice, because the sequence of terms can also describe problems users of these patterns try to resolve.

The softmax regression model by Zhang et al. [13] was used to compute the probabilities of using organizational patterns. The sorted probabilities from the output of the softmax regression model were used to establish sequences of organizational patterns expected to be used. Because Zhang et al. [13] noted it as possible, this softmax regression model was implemented as an artificial neural network using the Tensorflow framework.

The softmax regression model by Zhang et al. [13] was designed to produce soft predictions. In the output from this model, soft predictions meant probabilities that fragments of text descriptions of organizational patterns can be ascribed to the names of these patterns. A neural network with nine hidden processing layers was used to work with term frequencies - inverse document frequencies (TF-IDF) of pairs of words used in text descriptions of organizational patterns, and four hidden layers were used to work with TF-IDF of triplets of words. Softmax regression model was applied in two experiments with TF-IDF of pairs of words, and in two experiments with TF-IDF of triplets of words in text descriptions of organizational patterns from Coplien and Harrison [5].

5.0.1 Evaluation

The meaningfulness of the four established pattern sequences was successfully evaluated by describing them in pattern stories.

Classification accuracy of the artificial neural network used in both experiments was evaluated using confusion matrices on training, validation, and test sets to determine if it is sufficiently strong to produce probabilities that can be used to identify patterns expected to be used together in pattern sequences. The inability of neural networks to classify well resulted from the similarity of input TF-IDF frequencies.

Sensitivity analysis was conducted on the artificial neural network used in experiments with TF-IDF of triplets of words to evaluate its sensitivity to errors in input frequencies. The neural network showed no sensitivity.

5.0.2 Results

Four meaningful pattern sequences of organizational patterns expected to be used were established during two experiments with TF-IDF of pairs of words in text descriptions of organizational patterns. Four meaningful pattern sequences of organizational patterns expected to be used were established during two experiments with TF-IDF of triplets of words in text descriptions of organizational patterns.

6. Conclusions and Future Work

Relationships between software patterns that lead to the identification of meaningful pattern sequences can be discovered from their descriptions using probabilistic methods. Some of these relationships lead to the identification of pattern sequences that are expected to be used in practice. This thesis presented various stochastic processes that can be used to identify meaningful pattern sequences. The use of these stochastic processes supported the thesis by identifying implicit relationships between software patterns that could be used to establish meaningful pattern sequences. It was shown that these implicit relationships also lead to identifying pattern sequences that are expected to be used in practice. It was shown that sequences of patterns that are expected to be used in practice can solve complex problems in software development and can be used in establishing security in organizations. These pattern sequences can also help pattern users to orient themselves in many patterns if they lack knowledge on which patterns are expected to be used together.

The thesis presented three methods of establishing sequences of organizational and security patterns using implicit relationships extracted from text descriptions of organizational and security patterns to identify new, previously unknown pattern sequences and already seen pattern sequences as innovative solutions to organizational and security problems.

The advantage of using the methods presented in this thesis lies in their ability to produce meaningful pattern sequences based on stochastic relationships and the ability to establish sequences of patterns from various pattern languages, pattern catalogs, or application domains. Another advantage of using these methods employing stochastic relationships is that these methods do not require statistics about the past use of patterns to establish pattern sequences, because the massive use of patterns cannot be observed. Methods presented here:

- Establish new sequences of organizational and security patterns. These methods can also be used for other types of software patterns.
- Establish sequences of organizational and security patterns that are the most expected to be used, expected to be used, and unexpected to be used.
- Establish sequences of organizational and security patterns, which can be combined to establish security in organizations.
- Establish pattern sequences that are most expected to be used that stay meaningful even if the order of their subsequently used patterns is exchanged.
- Establish meaningful pattern sequences without using statistics about the past use of software patterns in practice.
- Two methods presented in the Chapter 4 can be used to establish pattern sequences from text descriptions of patterns which link to each other, or are otherwise implicitly related.
- The method presented in Chapter 5 can be used to establish pattern sequences from text descriptions of software patterns, which do not have to link to each other. This

method can be used to establish pattern sequences from patterns from multiple application domains without having to execute this method multiple times (as in the case for methods presented in Chapter 4.

The method of establishing pattern sequences using stochastic trees and the method of establishing pattern sequences using Bayesian networks should be tested with probabilities calculated using the Beta distribution, thus supporting pattern users to insert their own experiences with the use of software patterns. Both methods should also be evaluated on design patterns to see the impact of established pattern sequences on the architecture of software systems. Other structures of Bayesian networks can be tried to determine the best structure and parameters explaining the expected establishment of pattern sequences. The structure and parameters of Bayesian networks should also be learned from text descriptions of software patterns, or their suitable representation of relationships. These structures can then be queried using dynamic programming languages to answer multiple questions about the expectancy of pattern sequences simultaneously.

The performance of the Softmax regression model must be compared to other architectures, such as Recurrent neural networks.

Bibliography

[1] C. Alexander, S. Ishikawa, and M. Silverstein. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, New York, August 1977. ISBN 0195019199.

[2] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, Cambridge, UK, 04-2011 edition, 2011. In press.

[3] S. D. Buschmann F., Henney K. *Pattern Oriented Software Architecture: On Patterns and Pattern Languages*. Wiley, 2007.

[4] J. Coplien and L. Zhao. Symmetry breaking in software patterns, 10 2000.

[5] J. O. Coplien and N. B. Harrison. Organizational patterns of agile software development. *Science*, 2004.

[6] C. Cordeiro, Vasconcelos. A catalog of security patterns, 2022.

[7] G. Hazen. Stochastic trees: A new technique for temporal medical decision modeling. *Medical decision making : an international journal of the Society for Medical Decision Making*, 12:163–78, 08 1992. doi: 10.1177/0272989X9201200302.

[8] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[9] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009. ISBN 0262013193.

[10] J. Kruschke. *Doing Bayesian Data Analysis (Second Edition)*. Academic Press, Boston, 2015.

[11] T. B. Sousa, H. S. Ferreira, and F. F. Correia. A survey on the adoption of patterns for engineering software for the cloud. *IEEE Transactions on Software Engineering*, 48(6):2128–2140, 2022. doi: 10.1109/TSE.2021.3052177.

[12] S. Waseeb, W. S. Khail, H. G. Wahaj, and V. Vranić. Extracting relations between organizational patterns using association mining. In *Proceedings of the European Conference on Pattern Languages of Programs 2020, EuroPLoP '20*, pages 1–9, 2020.

[13] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.

A. Publications

A.1 Articles in International Scientific Journals

V. Matovič (80%) and V. Vranić. Establishing Pattern Sequences Using Stochastic Trees and Bayesian Belief Networks With an Application to Security Patterns. IEEE Access, 13: 32443–32457, 2025. IEEE. doi: 10.1109/ACCESS.2025.3543319. (SJR Q1 / JCR Q2)

V. Matovič (70%), V. Vranić, W. Sulaiman Khail, and N. Harrison. Establishing Pattern Sequences Using Stochastic Processes with an Application to Organizational Patterns. Submitted to Springer Nature Social Network Analysis and Mining. Under review. (SJR Q1)

A.2 Articles in Proceedings of International Scientific Conferences

V. Matovič (90%) and V. Vranić. 2024. Establishing Pattern Sequences Using Stochastic Processes with an Application to Security Patterns. In Proceedings of the 29th European Conference on Pattern Languages of Programs, People, and Practices, EuroPLoP 2024. Kloster Irsee in Bavaria, Germany. ACM. doi: 10.1145/3698322.3698357. (Scopus)

V. Matovič (90%) and Valentino Vranić. Establishing Pattern Sequences Using Artificial Neural Networks with an Application to Organizational Patterns. In Proceedings of 10th Workshop on Software Quality Analysis, Monitoring, Improvement, and Applications, SQAMIA 2023. Bratislava, Slovakia. CEUR Workshop Proceedings, 2023. (Scopus)